

Theoretische Untersuchungen zur MHC I Antigenpräsentation

Dissertation

Zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Biophysik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I

der Humboldt-Universität zu Berlin

von

Diplom-Biophysiker Sascha Bulik

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Andreas Herrmann

Gutachter: 1. Prof. Dr. Hermann-Georg Holzhütter

2. Prof. Dr. Hanspeter Herzel

3. Prof. Dr. Burkhardt Dahlmann

Tag der mündlichen Prüfung: 10.06.2011

Zusammenfassung

Der MHC I Pathway ist ein Teil des Immunsystems und stellt mittels Antigenen an der Zelloberfläche den Proteinstatus der Körperzellen dar. Ziel dieser Arbeit ist durch die Entwicklung von Modellen zu Proteinsynthese und Abbau sowie den Teilschritten des MHC I Pathways und der Untersuchung von Simulationsergebnissen das Verständnis der zugrunde liegenden Prozesse zu verbessern und gegebenenfalls die Qualität der Antigenprädiktion zu verbessern.

Es wurden statistische Modelle für den Transport der Peptide in das ER mittels TAP, das Schneiden von Peptidbindungen durch das Proteasom und das cytosolische sowie endoplasmatische Trimmen von Peptiden entwickelt. Weiterhin wurden kinetische Modelle zur Synthese und Abbau von viralen Proteinkonstrukten, zur proteasomalen Erstellung von Proteinfragmenten und zum Simulieren des Gesamtprozesses von Infektion bis zur Antigenpräsentation entwickelt.

Es wurde gezeigt, dass eine DRiP Rate von 10 Prozent die Antigenpräsentation unabhängig von der Lebenszeit der Proteine gewährleistet und für langlebige Proteine der Anteil der Antigene aus DRiPs die Gesamtmenge der Antigene dominiert. So wird gewährleistet, dass an der Zelloberfläche der Status der momentanen Proteinsynthese dargestellt wird.

Die erstellten Teilmodelle der Schritte des MHC I Pathways sind jeweils auf *in vitro* Daten des jeweiligen Prozesses trainiert und ermöglichen zusammen mindestens die gleiche Vorhersagequalität, wie Pathway Modelle, die auf *in vivo* Daten trainiert worden sind. Dies zeigt, dass alle wesentlichen Prozesse zur Antigenpräsentation von den erstellten Modulen erfasst werden. Die Module können je nach Bedarf und Fragestellung zu Modellen kombiniert werden. Ein Vorhersagetool wurde auf <http://mhc-pathway.net> zur Verfügung gestellt.

Durch Modellanalysen können der relative Beitrag der einzelnen Schritte des Pathways und die Voraussetzungen für ein potentiell Antigen bestimmt werden.

Das kinetische Modell der Prozesse von Infektion bis Antigenpräsentation erlaubt das Verfolgen aller Peptide und das Analysieren der Prozesse die zur Erstellung von Epitopen führen. Die quantitativen Vorhersagen können experimentell validiert werden.

Die proteasomale Fragmenterstellung ist der Teilprozess, der noch am wenigsten gut verstanden ist und bedarf noch weiterer experimenteller Untersuchungen.

Schlagwörter:

Modellierung, MHC-I Antigenpräsentation, Immunologie, Defektive ribosomale Produkte, TAP, Proteasom

Abstract

The MHC I antigen presentation pathway is part of the immune system and enables cells to show their proteome state at the cell surface. This work aims at improving the understanding of the MHC I pathway and the prediction of antigens from the source proteins where appropriate. The means are the development of models for protein synthesis, degradation and the individual steps of the pathway as well as the analysis of simulation results.

Statistical models for the transport of peptides into the ER by TAP, the cleavage of peptide bonds by the proteasome, and the cytosolic and endoplasmic trimming of peptides have been developed. Furthermore, kinetic models for synthesis and degradation of viral protein constructs, for proteasomal generation of protein fragments, and for the simulation of the entire process from viral infection of a cell to the resulting antigen presentation were created.

It has been shown that a DRiP rate of 10% is sufficient to have antigen presentation independent of the source protein's live time and that the antigens derived from the DRiP pool dominate the antigen presentation for long lived proteins. This mechanism enables the presentation of the current protein synthesis state of the cell.

Each developed model of a part of the MHC I pathway is trained on in vitro data and together they provide at least the same prediction quality as pathway models that are based on in vivo data. This shows that all processes that contribute significantly to antigen presentation are covered in the developed models. The models for the individual parts can be combined according to the demand and question. A prediction tool has been provided at <http://mhc-pathway.net>.

The contribution of each part of the pathway can be assessed by model supported analysis of the individual steps. It is possible to determine the traits of a potential antigen.

The kinetic model of the pathway from infection to antigen presentation enables the generation of time curves for each possible peptide and the analysis of the processes that lead to the development of antigens. The quantitative predictions allow for experimental validation.

The proteasomal fragment generation is the least good understood part of the MHC I pathway and requires further experimental study.

Keywords:

Modelling, MHC-I antigen presentation, Immunology, Defective ribosomal products, TAP, Proteasome

Inhaltsverzeichnis

Zusammenfassung	I
Abstract.....	III
Inhaltsverzeichnis	V
1 Einführung	1
1.1 Der MHC I Präsentationsweg	1
1.1.1 Major Histocompatibility Complex (MHC)	1
1.2 Zelluläre Proteinsynthese und Ubiquitin gesteuerte Proteolyse	3
1.2.1 Abbau von neu synthetisierten Proteinen – die DRiP Hypothese	4
1.3 Die Komponenten des MHC Class I Antigen Präsentationsweges	5
1.3.1 Das Proteasom	5
1.3.2 Der Transporter assoziiert mit Antigen Präsentation – TAP.....	7
1.3.3 Das MHC I Molekül	8
1.3.4 Cytosolische und endoplasmatische Peptidasen	10
1.3.5 Ausnahmen vom klassischen MHC I Pathway.....	11
1.4 Aufgabenstellung	11
2 Der Beitrag von DRiPs zur Antigenpräsentation	13
2.1 Hintergrund und experimentelle Daten	13
2.1.1 Experimenteller Aufbau.....	13
2.1.2 Experimentelle Ergebnisse	14
2.2 Modell und Methoden	15
2.2.1 Das kinetische Modell	15
2.2.2 Anpassen der Modellparameter an die experimentellen Daten.....	19
2.3 Ergebnisse	20
2.3.1 Nur einige experimentelle Kurven lassen sich durch das Modell reproduzieren.....	20
2.3.2 Verschiedene Modellvariationen wurden getestet.	21
2.3.3 Hemmung der Proteinsynthese verringert die proteasomale Abbaurrate.....	24
2.3.4 Auswertung der Modellparameter	25
2.3.5 Der Einfluss von DRiPs auf die Antigenpräsentation.....	27
2.4 Diskussion.....	30
2.4.1 Eine DRiP Rate von 10% ist ausreichend für die Antigenpräsentation.....	31
2.4.2 DRiPs sind nützlich für die Immunüberwachung und beschleunigen die Präsentation metabolisch stabiler Proteine.	33

2.4.3	Das Verhältnis von abgebauten Proteinen zu präsentierten Antigenen ist im Mittel 1'000 zu 1 und ist wahrscheinlich gleich für DRiPs und native Proteine.	33
2.4.4	Hinweise auf eine wechselseitige Koordination von Proteinsynthese und Proteinabbau.....	34
2.4.5	Ist das Modell zu stark vereinfacht, um verlässliche Aussagen über die Rolle der DRiPs in der Antigenpräsentation abzuleiten?	36
3	Vorhersage von potentiellen Epitopen aus der Proteinsequenz.....	37
3.1	Peptid Bindung an MHC Class I Moleküle	38
3.1.1	Verwendete Modelle für MHC Class I Allele.....	38
3.2	TAP Transport	39
3.2.1	Verwendete experimentelle Daten	39
3.2.2	Beschreibung der Peptidaffinität zu TAP.....	43
3.2.3	Erweiterung des TAP Modells mit Paarkoeffizienten	48
3.2.4	Erweitern des TAP Modells für längere Peptide.....	50
3.2.5	Einfluss von TAP auf die Antigenpräsentation	52
3.2.6	Diskussion des humanen TAP Modells und des Einflusses von TAP auf die Antigenpräsentation	61
3.3	Proteasomale Schnittpräferenz	63
3.3.1	Verwendete experimentelle Daten	63
3.3.2	Methode der Modellentwicklung	65
3.3.3	Ergebnisse.....	66
3.3.4	Diskussion	73
3.4	N-terminales Trimmen von Peptiden.....	76
3.4.1	Verwendete Daten und Methoden	76
3.4.2	Ergebnisse.....	79
3.4.3	Diskussion	83
3.5	Kombination verschiedener Schritte des MHC I Pathways	86
3.5.1	Sequentielles Verwenden von einem Modell als Filter für das nächste Modell ist bei geeignetem Cutoff möglich.....	86
3.5.2	Die Kombination von zwei Schritten des Pathways mit der Filter Methode verbessert die Vorhersagequalität.....	86
3.5.3	Durch additive Kombination der Modelle erhält man relative Peptidmengen.	89
3.5.4	Kombination der Modelle mit N-terminaler Trimmung von Peptiden führt zu keiner Verbesserung der Vorhersagequalität.....	94
3.5.5	Diskussion	95

4	Mechanistisches Modell der proteasomalen Fragmentgeneration	101
4.1	Entwicklung des „slide & slice“ Modells	101
4.1.1	Das Erzeugen der richtigen Längenverteilung	101
4.1.2	Beschreibung des „slide & slice“ Modells	102
4.1.3	Anpassung der Modellparameter an experimentelle Daten	103
4.2	Ergebnis	104
4.2.1	Das Modell gibt die experimentell beobachtete Längenverteilung gut wieder	104
4.2.2	Fragmente lassen sich nicht gut vorhersagen	107
4.2.3	Einbeziehen von Sekundärstrukturinformationen verbessert die Vorhersage nicht	109
4.3	Diskussion	111
5	Kinetische Modellierung der Epitop Generierung	113
5.1	Antigenprozessierung aus pp89-25mer	113
5.1.1	Beschreibung des verwendeten Modells	113
5.1.2	Ergebnisse der Simulation	116
5.1.3	Diskussion	120
5.2	Antigenprozessierung des HIV-1-BRU gag Proteins	121
5.2.1	Charakterisierung des Proteins und des Modells	122
5.2.2	Ergebnisse der Simulation	124
5.2.3	Diskussion	126
6	Zusammenfassung der Ergebnisse und Schlussfolgerungen	129
7	Ausblick	135
	Literatur	137
	Publikationen	149
8	Anhang	151
8.1	Die stabilisierte Matrix Methode (SMM)[110]	151
8.2	Weitere Methoden zur Vorhersage von Peptideigenschaften (Übersicht)	152
8.2.1	Direktes Messen der Affinität mittels Peptidbibliotheken	152
8.2.2	Das Einbeziehen von Paarkoeffizienten	153
8.2.3	Künstliche neuronale Netze, Hidden Markov Modells, Regression Trees und Random Forests	154
8.2.4	Supervised learning	156
	Erklärung zur Selbstständigkeit	159

1 Einführung

1.1 Der MHC I Präsentationsweg

Das Immunsystem versetzt Organismen in die Lage, Krankheitserreger wie Mikroorganismen, körperfremde Substanzen, Viren und entartete körpereigene Zellen abzuwehren und auf diese Art die Schädigung des restlichen Organismus zu verhindern.

1.1.1 Major Histocompatibility Complex (MHC¹)

Der MHC Class I Antigenpräsentationsweg ist Teil des Immunsystems und ermöglicht es Zellen, Fragmente aus in der Zelle synthetisierten Proteinen an der Zelloberfläche zu präsentieren[1, 2, 3, 4]. Er umfasst alle Prozesse von der Erzeugung von Proteinfragmenten bis zur Präsentation an der Zelloberfläche.

Cytosolische Proteine werden durch das Ubiquitin-Proteasom-System abgebaut und in kurze Fragmente gespalten. Während der größte Teil dieser Peptide durch Aminopeptidasen in die einzelnen Aminosäuren hydrolysiert wird, transportiert der Transporter assoziiert mit Antigenpräsentation (TAP) einen kleineren Teil in das endoplasmatische Retikulum (ER). Dort hydrolysieren wiederum Aminopeptidasen die Fragmente, aber einige Fragmente binden an MHC Moleküle und bilden MHC Komplexe. Das Protein Tapasin unterstützt dabei den Komplexbildungsprozess, indem es MHC Moleküle in die Nähe von des TAP Transporters bringt. Die MHC Komplexe werden mittels vesikulären Transports über den Golgi Apparat an die Zelloberfläche gebracht, wo die Antigene präsentiert werden und von T-Zellen mit ihren Rezeptoren erkannt werden können.

¹ MHC **M**ajor **H**istocompatibility **C**omplex; Haupthistokompatibilitätskomplex bezeichnet eine Gruppe von Genen und auch Proteinen, die für die Immunerkennung und die Gewebeverträglichkeit bei Transplantationen wichtig sind.

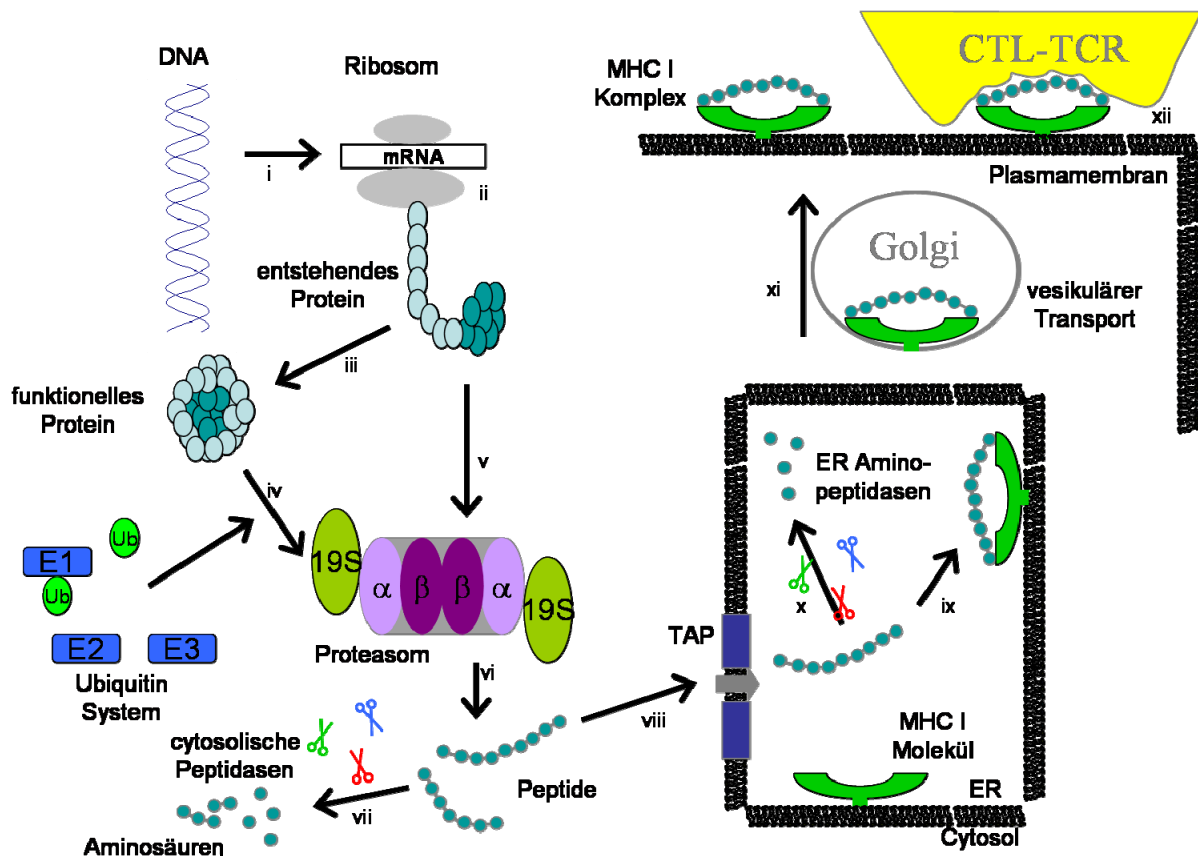


Abbildung 1: Überblick über den MHC Class I Pathway

i: Transkription der DNA zur mRNA; mRNA Spleißen und andere Interaktionen auf mRNA Level. ii: Translation der mRNA am Ribosom. iii: Faltung und posttranslationale Modifikation des entstehenden Proteins zum funktionellen Zustand. iv: Abbau der funktionellen Proteine z. T. mit Hilfe des Ubiquitin Systems durch das Proteasom. v: Abbau von gerade entstehenden Proteinen oder defekten ribosomalen Produkten (DRiPs; auch z. T. mit Hilfe des Ubiquitin Systems) durch das Proteasom. vi: Erstellung von Peptidfragmenten durch das Proteasom. vii: Hydrolytische Spaltung der Peptide im Cytosol in ihre Aminosäuren durch cytosolische Peptidasen. viii: Bindung von cytosolischen Peptiden an TAP und Transport der Peptide in das ER. ix: Bindung endoplasmatischer Peptide an MHC I Moleküle und Bildung von MHC I Komplexen. x: Hydrolyse von Peptiden im ER durch endoplasmatische Amino-peptidasen, Export der Fragmente in das Cytosol mittels Sec61. xi: vesikulärer Transport der MHC I Komplexe zur Zelloberfläche über den Golgi Apparat. xii: Präsentation der MHC I Komplexe an der Zelloberfläche und Bindung durch cytotoxische T-Zellen mit dem entsprechenden Rezeptor.

Das konzertierte Zusammenspiel der Komponenten des Pathways ist das Thema dieser Arbeit, daher werden die einzelnen Komponenten des Systems detailliert in Abschnitt 1.3 erläutert.

Beim Menschen befindet sich die MHC Region auf Chromosom 6 und umfasst mehr als 200 Gene, von denen etwa die Hälfte als exprimiert vorhergesagt wird. Viele dieser Gene kodieren für Proteine, die Teil des MHC Systems sind, wie die Untereinheiten des TAP Transporters, die MHC I Moleküle, die MHC II Moleküle oder die Immununtereinheiten des Proteasoms[5, 6].

Das humane Leukozytenantigen-System (HLA)[7]

Jeder Mensch hat jeweils drei Genorte, die für MHC I Moleküle und für MHC II Moleküle kodieren. Aufgrund des diploiden Chromosomensatzes exprimiert somit jeder Mensch jeweils sechs verschiedene MHC Moleküle für die beiden Klassen. Die Genorte für MHC I werden als HLA-A, HLA-B und HLA-C bezeichnet. Die Genorte für MHC II heißen HLA-DR, HLA-DP und HLA-DQ. Es sind mehr als 4000 spezifische Varianten und Subtypen für MHC I und mehr 1000 Varianten und Subtypen für MHC II bekannt[8].

Das Immunsystem muß zwischen gesunden und geschädigten beziehungsweise infizierten Zellen zu unterscheiden, um den Organismus vor der Ausbreitung des schadhafte Prozesses zu schützen. In dieser Arbeit wird der MHC Class I Antigen Präsentationsweg untersucht und seine Bestandteile modelliert.

Auf der Basis des Verständnisses des MHC Class I Pathways könnten beispielsweise Impfstoffe mit geringeren Nebenwirkungen hergestellt und Autoimmun- und Tumorerkrankungen besser verstanden werden.

1.2 Zelluläre Proteinsynthese und Ubiquitin gesteuerte Proteolyse

Im MHC I Antigenpräsentationsweg werden Peptide aus von der Zelle synthetisierten Proteinen an der Zelloberfläche dargestellt.

Die in der DNA enthaltene genetische Information wird in mRNA transkribiert Abbildung 1 i). Die mRNA wird durch verschiedene Prozesse modifiziert, bevor sie am Ribosom zu Proteinen translatiert wird (Abbildung 1 ii), so werden gegebenenfalls Introns entfernt und das 3'Ende polyadenyliert, um die Stabilität gegenüber Ribonukleasen zu erhöhen. Die Translationseffizienz ist zum einen eine intrinsische Eigenschaft der spezifischen mRNA, wird aber auch durch RNA-Interferenz z. B. durch siRNA reguliert[9, 10]. Die mRNA wird durch das Exosom abgebaut[11, 12]. Am Ribosom wird die mRNA schrittweise translatiert, wobei Aminosäuren von Aminoacyl-tRNAs einzeln zu der dabei entstehenden Polypeptidkette hinzugefügt werden[13]. Die Polypeptidkette muss sich zum funktionellen Protein falten wobei zum großen Teil Chaperone[14, 15] behilflich sind. Auch weitere Pro-

zesse, wie die Bildung von Disulfidbindungen und posttranslationale Modifikation von Aminosäuren sind dafür notwendig, dass das Protein den endgültig funktionalen Zustand einnimmt (Abbildung 1 iii).

Der Bedarf der Zelle, welche Proteine zu einem Zeitpunkt vorhanden sein müssen, ändert sich mit den momentanen Aufgaben. Proteine, die nicht mehr benötigt werden aber auch Proteine, die durch chemische Modifikationen unbrauchbar geworden sind, werden abgebaut. Durch das Ubiquitin-Ligase System[16, 17] können Proteine gezielt für den Abbau durch das Proteasom markiert werden (Abbildung 1 iv).

Die Lebenszeit von Proteinen, die nicht zu einem speziellen Zeitpunkt auf ein Signal hin abgebaut werden, ist eine inherente Eigenschaft des Proteins und es ist bisher nicht geklärt, was die metabolische Stabilität eines Proteins bestimmt. Die Änderung der N-terminalen Aminosäure eines Proteins hat einen starken Einfluss auf seine Stabilität[18]. In neuerer Zeit werden die Lebenszeiten von vielen Proteinen experimentell bestimmt[19, 20].

1.2.1 Abbau von neu synthetisierten Proteinen – die DRiP Hypothese

Ein Teil der am Ribosom neu synthetisierten Proteine wird direkt vom Proteasom abgebaut und trägt zur Antigenpräsentation bei ohne einen funktionellen Zustand einzunehmen (Abbildung 1 v)[21, 22].

Es wurde gezeigt, dass die Neusynthese von sehr langlebigen Proteinen notwendig ist, um Antigene aus diesen Proteinen an der Zelloberfläche beobachten zu können[23]. Obwohl diese Proteine noch in den Zellen nachzuweisen sind, reicht ihr Abbau nicht aus, um genügend Antigene zu produzieren, dass sie mit T-Zellen detektiert werden können. Auch der starke Rückgang von Peptidtransporten in das ER mittels TAP[24] bei Inhibition der Proteinsynthese zeigt, dass ein Großteil der Antigene aus neu synthetisierten Proteinen stammt.

Es ist nicht bekannt, ob es sich bei diesem Pool um Proteine handelt, die in ihrer Sequenz fehlerhaft sind, ob sie noch während der Translation abgebaut werden oder ob sie es nicht schaffen, sich zu funktionellen Proteinen zu falten, bevor sie als fehlgefaltet erkannt und für den Abbau markiert werden. So zeigten Turner et al., dass ein Teil der Proteine mit starkem N-terminalem Degradationsmotiv[18] noch während der Proteinsynthese abgebaut wird[25]. Während in einzelnen Fällen auch untypische Peptide, z.B. aus nicht codierenden Abschnitten der DNA[26], präsentiert werden, ist es schwer vorstellbar, dass die Antigenpräsentation zu einem großen Teil von Peptiden abhängig ist, die in ihrer Sequenz fehlerhaft sind. Dieses widerspricht der Vorstellung eines hochsensitiven Systems

der Immunüberwachung. Das Cytosol der Zellen hat eine hohe Konzentration an Proteinen und anderen Molekülen, so steht neu synthetisierten Proteinen nur eine kurze Zeit zur Verfügung, in der sie sich zu funktionellen Proteinen falten können, ohne mit anderen Strukturen zu aggregieren[27]. Obwohl der Prozess der Faltung durch Chaperone unterstützt wird[14, 15, 28], steht er in Konkurrenz zu Abbauprozessen.

Der Pool der schnell abgebauten Proteine wird als „Defective Ribosomal Products“ (DRiPs) bezeichnet[21, 29]. Der Anteil dieses Pools an der MHC I Antigenpräsentation wird in der Literatur mit mehr als 1/3 abgeschätzt[21, 30, 31].

1.3 Die Komponenten des MHC Class I Antigen Präsentationsweges

1.3.1 Das Proteasom

Im klassischen MHC Class I Pathway[1, 2, 3, 4]² werden cytosolische Proteine durch das Proteasom in kleine Fragmente, mit Längen zwischen 4 und 30 Aminosäuren, zerlegt (Abbildung 1 vi). Substrate sind vor allem ubiquitinierte Proteine. Es wird ebenfalls vermutet, dass das 26S Proteasom zum Teil zunächst relativ lange Fragmente erstellt, die dann durch Aminopeptidasen und TPPII zu kürzeren Peptiden hydrolysiert werden, dann durch das 20 S Proteasom in die endgültigen proteasomalen Produkte zerlegt werden. Für den zweiten proteasomalen Schritt ist keine Ubiquitinierung der Substrate notwendig[32].

Das Bereitstellen von Peptiden als Antigene ist nicht die einzige Aufgabe des Proteasoms. Die Hauptaufgabe des Proteasoms ist der Abbau von Proteinen, wodurch die enthaltenen Aminosäuren der Zelle wieder zur Synthese neuer Proteine zur Verfügung stehen. Dieses wird durch das zum Überleben notwendige Vorhandensein von Proteasomen in fast allen Zellen, also auch Einzellern ohne Immunsystem belegt. In höheren Organismen ist das Bereitstellen von Antigenen zur Präsentation ebenfalls eine wichtige Funktion des Proteasoms.

Das Proteasom ist ein großer Proteinkomplex, der aus zahlreichen Untereinheiten aufgebaut ist. Es besteht zum einen aus dem proteolytisch aktivem Kern, dem 20 S Proteasom, das wiederum aus vier Ringen aufgebaut ist. Das 20 S Proteasom ist in Abbildung 2 dargestellt. Die äußeren Ringe bestehen jeweils aus sieben verschiedenen α - und die inneren aus sieben verschiedenen β -Untereinheiten, von denen drei Proteaseaktivität (tryp-

² Soweit nicht anders angegeben, stammen die im Kapitel 1.3 enthaltenen Informationen aus diesen Reviews.

tisch, chymotryptisch und caspaseähnlich) aufweisen. Beim Übergang vom konstitutiven zum Immunoproteasom durch Interferon Gamma (IFN- γ) werden die drei aktiven Untereinheiten ausgetauscht. Die für die Immununtereinheiten codierenden Gene werden hierbei induziert und es kommt so zur Synthese der Immununtereinheiten, die dann beim Zusammenbau neuer Proteasome als Bausteine zur Verfügung stehen. Der Einbau von konstitutiven Untereinheiten in den Proteasomkomplex wird dabei reduziert.

Zum anderen besteht das Proteasom auch noch aus regulierenden Untereinheiten, dem PA28 (11 S) oder dem PA 700 (19 S) Komplex, die aus mehreren Untereinheiten aufgebaut sind. Die regulierenden Untereinheiten binden an den α -Ring des Proteasoms. Der PA 700 Komplex weist ATPaseaktivität auf und PA 28 spielt möglicherweise eine Rolle in der Herstellung von Antigenkandidaten, da dieser Komplex ebenfalls durch IFN- γ induziert wird.

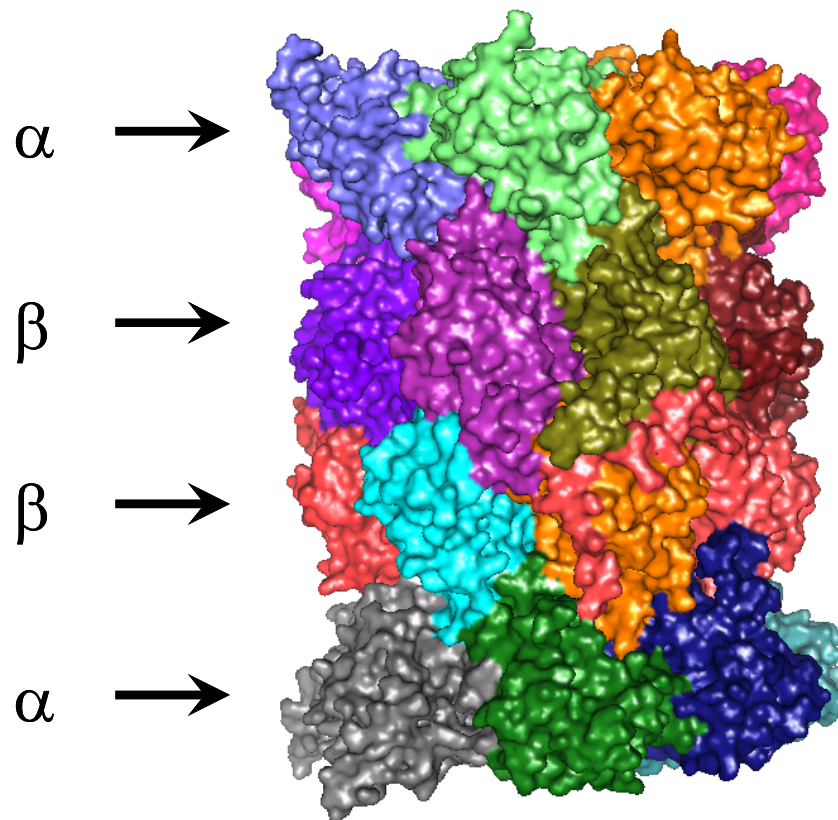


Abbildung 2: 20S Proteasom, Struktur aus Groll et al. [33]

Das 20S Proteasom besteht aus 28 Untereinheiten, oben und unten jeweils ein α -Ring in der Mitte zwei β -Ringe mit den katalytischen Zentren. Jeder Ring besteht aus sieben Untereinheiten, die nicht alle sichtbar sind, unterschiedliche Untereinheiten sind in unterschiedlicher Farbe gekennzeichnet. Jeweils drei aktive Untereinheiten sind in den β -Ringen.

1.3.2 Der Transporter assoziiert mit Antigen Präsentation – TAP

Die meisten vom Proteasom erzeugten Peptide werden durch cytosolische Peptidasen in Aminosäuren zerlegt und stehen so der Zelle wieder für den normalen Stoffwechsel zur Verfügung. Einige Fragmente werden jedoch mittels TAP unter Verbrauch von ATP in das ER transportiert[34, 35, 36] (Abbildung 1 viii).

TAP ist ein ABC Transporter[37, 38] vom Typ B, also eine Heterodimer. Jedes Monomer besteht aus einer Nukleotid bindenden Domäne (NBD) und einer Transmembrandomäne (TMD) und wird als TAP1 bzw. TAP2 bezeichnet. Die Monomere sind einzeln nicht aktiv, es kommen auch keine funktionstüchtigen TAP1/TAP1 beziehungsweise TAP2/TAP2 Dimere vor. Es existiert eine Röntgenkristallstruktur der NBD von TAP1[39] und außerdem elektronenmikroskopische Aufnahmen des Dimers[40]. Pro TMD durchspannen sechs α -Helices die Membran des ER. Die NBD haben die für ABC Transporter üblichen Walker AB Motive zur Bindung von ATP und den charakteristischen C- Loop.

Es gibt auch Untersuchungen zur Bindungsstelle für die Peptide, aber bisher ist nur gezeigt worden, dass die Mutation einzelner spezieller Aminosäuren in den NBD eine Änderung des Bindungsverhaltens mit sich führt[41, 42, 43]. Die exakte Region ist auf Strukturebene noch nicht charakterisiert worden. Die Annahme, dass mehrere Bindungsregionen für Peptide existieren, ist durch Untersuchungen mit fluoreszenzmarkierten Peptiden und mit Photocrosslinkern widerlegt worden[44]. Es wird vermutet, dass sich die Bindungsregion über Teile beider NBD von TAP erstreckt[45].

Die Gene für den TAP Transporter sind beim Menschen auf dem Chromosom 6 im MHC Class II Locus bestimmt worden[46]. Obwohl die Gene im MHC Class II Locus sind, werden sie mit den anderen MHC Class I Genen reguliert. Die TAP Gene bestehen für jedes Monomer aus etwa 10 Kilobasen. Die Genexpression wird durch Cytokine reguliert, so führt IFN- γ zu einer erhöhten Expression[47] und Interleukin 10 zu einer Reduktion der Genexpression. Es existieren verschiedene polymorphe Formen der menschlichen TAP Monomere. Für TAP1 existieren drei verschiedene Typen und für TAP2 fünf. Die Literatur ist nicht einheitlich, aber mehrheitlich wird die Ansicht vertreten, dass es keinen funktionellen Unterschied zwischen den Allelen gibt[48, 49].

TAP transportiert ein breites Spektrum an Peptiden. Es werden Peptide, mit Längen zwischen sieben und 40 Aminosäuren, transportiert und auch künstlich erzeugte Peptide mit verzweigenden Seitenketten von mehreren Aminosäuren Länge. Die Affinität für die einzelnen Peptide unterscheidet sich jedoch sehr stark. So werden Peptide mit einer Länge von 8 bis 16 Aminosäuren mit hoher und annähernd gleicher Affinität gebunden, kürzere und längere Peptide jedoch nur mit geringer Affinität[50].

Einige Viren haben einen Weg gefunden, TAP zu blockieren und so der Immunabwehr des Körpers zu entgehen. Zu ihnen gehört der Herpes simplex Virus, das ein Protein ICP47 erzeugt, das als Inhibitor an TAP bindet[51, 52]. Es ist ermittelt worden, dass TAP durch Phosphorylierung inaktiviert wird und so von der Zelle reguliert wird [53].

1.3.3 Das MHC I Molekül

Aus dem Peptidpool im ER können einige Peptide hochspezifisch an MHC Class I Moleküle binden und so einen stabilen Komplex bilden (Abbildung 1 ix).

Die MHC Moleküle bestehen aus einer 45 kDa (heavy chain; HC) und einer kleinen 12 kDa Untereinheit (light chain; β_2m) dem β_2 -Mikroglobulin. Die HC Untereinheit ist in der Membran des ER verankert und die beiden Untereinheiten bilden in Abwesenheit des Antigenpeptids ein instabiles Dimer. Diese unbeladenen Moleküle werden durch die Chaperone Calnexin und Calreticulin stabilisiert und sind über das Protein Tapasin an TAP gebunden[50, 54]. Bei Beladung mit einem Peptid stabilisiert sich dieser Komplex (HC, β_2m und Antigen) und dissoziiert von TAP ab. Diese Verknüpfung des MHC Moleküls mit TAP hat den Vorteil, dass keine unbeladenen MHC Moleküle an die Zelloberfläche gelangen und ermöglichen vielleicht eine effizientere Beladung der MHC Class I Moleküle mit Peptiden, da die durch Diffusion zurückzulegende Strecke für das Antigen kürzer ist.

Die Peptide müssen für die Bindung an die MHC Moleküle spezielle Eigenschaften aufweisen. So werden im allgemeinen Peptide mit Längen zwischen 8 und 11 Aminosäuren akzeptiert, wenn sie allelspezifische, oftmals terminale Ankermotive aufweisen. Der Besitz der Ankermotive ist jedoch weder eine notwendige noch eine hinreichende Bedingung, um an das MHC Molekül zu binden, obwohl die meisten bekannten Epitope solche Motive aufweisen. Durch Bindungsstudien von Peptiden an MHC Moleküle ist die Spezifität von einigen Allelen gut untersucht worden. Abbildung 3 zeigt ein mit Peptid beladenes MHC I Molekül.

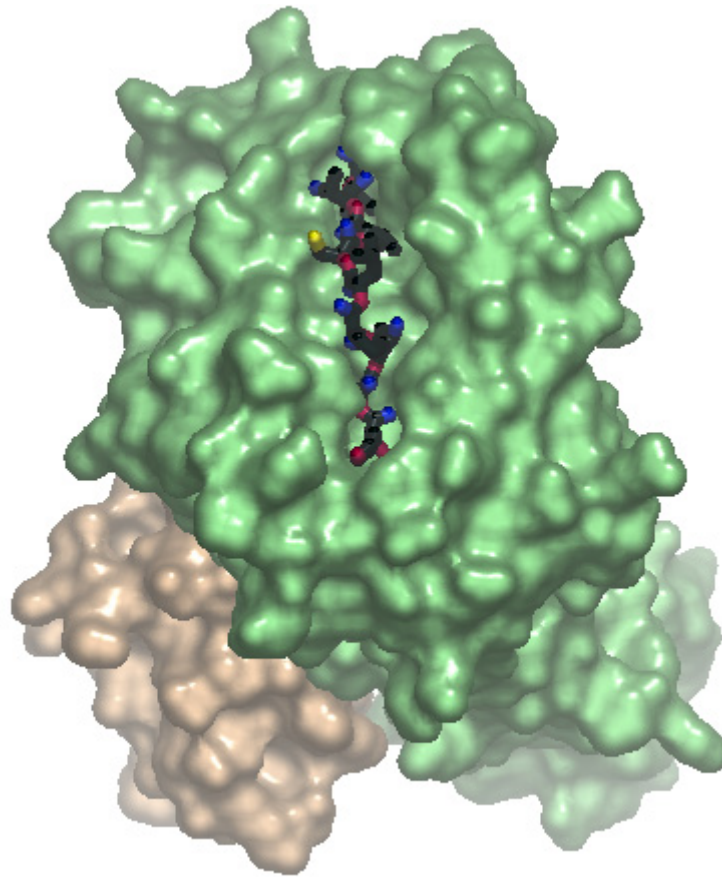


Abbildung 3: HLA-A2 Molekül mit gebundenem Peptid, Struktur aus Celie et al. [55]

Das Peptid ist im MHC Molekül eingespannt und kann von der Zellmembran weg orientiert.

Die MHC Komplexe werden aus dem ER über den Golgi Apparat in Vesikeln zur Zelloberfläche transportiert (Abbildung 1 xi), wo sie präsentiert werden. Dort können sie von cytotoxischen T-Zellen gebunden und so überprüft werden (Abbildung 1 xii). Antigene, die eine Immunantwort auslösen können, werden auch als T- Zell Epitope oder hier kurz als Epitope bezeichnet, in dieser Arbeit werden die beiden Begriffe Antigen und Epitop weitestgehend als Synonyme benutzt. In Abbildung 4 ist ein MHC I Komplex dargestellt, der von einem T-Zell Rezeptor gebunden wird.

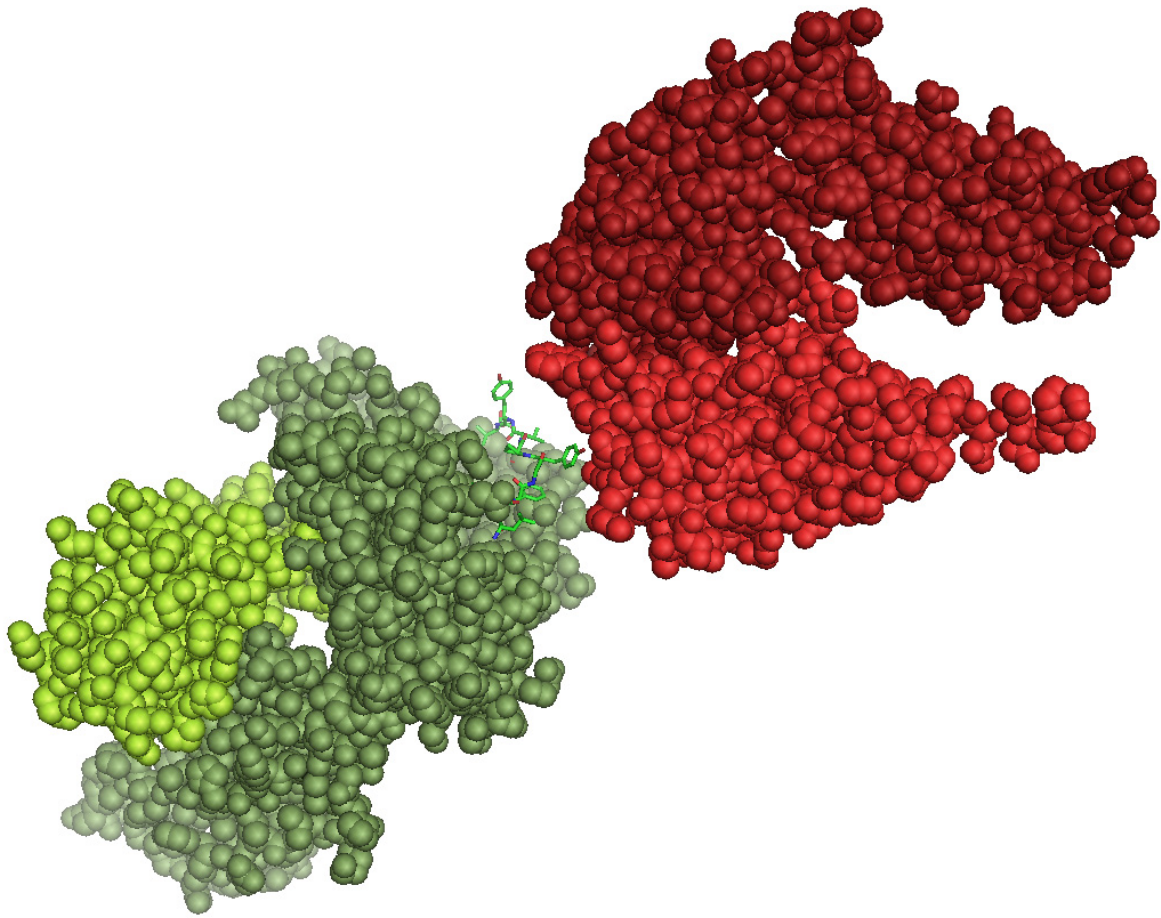


Abbildung 4: HLA-A2 Molekül mit Epitop und T-Zell Rezeptor, Struktur aus Garboczi et al. [56]

Das MHC Molekül ist grün dargestellt. Das Peptid ist eingespannt und wird präsentiert. Der Rezeptor der T- Zelle (rot) bindet das Peptid. Zur besseren Sichtbarkeit wurden die einzelnen Moleküle auseinandergerückt.

Nicht verwendete Peptide werden aus dem ER durch das ATP getriebene Translocon exportiert.

1.3.4 Cytosolische und endoplasmatische Peptidasen

Vom Proteasom erzeugte Peptide werden im Cytosol zum größten Teil in kürzere Peptide und einzelne Aminosäuren gespalten (Abbildung 1 vii), wobei der C-Terminus sehr stabil ist und der N-Terminus angegriffen wird. Einige Peptidasen sind auch als Endopeptidasen

aktiv. Die meisten Peptide werden noch im Cytosol hydrolysiert, für einzelne Peptide ist cytosolisches Trimmen³ eine notwendige Voraussetzung, um von TAP transportiert zu werden, während die Präsentationsrate einiger anderer Peptide durch die cytosolischen Peptidasen sinkt[57, 58, 59, 60, 61, 62].

Nach dem Transport der Peptide in das ER werden die Peptide dort von der endoplasmatischen Aminopeptidase ERAP 1 getrimmt und gegebenenfalls zerstört (Abbildung 1 x). Auch hier kann die Existenz der Aminopeptidasen notwendig für die Generierung des fertigen Antigens sein aber auch ein Konkurrenzprozess zur Bindung an das MHC I Molekül[3, 63, 64, 65].

1.3.5 Ausnahmen vom klassischen MHC I Pathway.

Es gibt eine ganze Reihe von Ausnahmen von diesem klassischen MHC Class I Pathway, aber die meisten Antigene gelangen auf diese Art zur Zelloberfläche[66, 67, 68, 69]. So ist beispielsweise gezeigt worden, dass Teile von Signalpeptiden als T-Zell Epitope präsentiert werden. Signalpeptide stammen von Proteinen, die kotranslational in das ER synthetisiert und im ER von der Signalase abgetrennt werden. Sie sind so unabhängig vom Proteasom und von TAP im ER verfügbar. Auch der Weg von Fragmenten aus abgebauten Mitochondrienproteinen ist noch nicht völlig klar. Andere Spekulationen, wie das Binden des Proteasoms an den TAP Transporter und somit das direkte Beliefern des ER mit Peptiden durch das Proteasom, existieren auch[70]. Weiterhin wurde gezeigt, dass die Tripeptidyl Peptidase II Antigene unabhängig vom Proteasom generieren kann.

1.4 Aufgabenstellung

Die genaue Kenntnis des Immunsystems und ins besondere des MHC I Präsentationsweges wird es ermöglichen Autoimmunkrankheiten zu bekämpfen und eine effizientere Vakzinierung gegen virale Krankheiten durchzuführen.

Das Verständnis der Prozessierung der Antigene ermöglicht es, Vakzine zu generieren, die effizienter präsentiert werden können.

In dieser Arbeit werden Modelle für die Komponenten des MHC I Antigenpräsentationsweges entwickelt und das konzertierte Zusammenspiel der beteiligten Prozesse, die zur Generierung eines Antigens an der Zelloberfläche führen, untersucht.

³ Trimmen: Abtrennen einer oder mehrer Aminosäuren vom N-Terminus eines Peptids

Im Kapitel 2 „Der Beitrag von DRiPs zur Antigenpräsentation“ dieser Arbeit wird ein quantitatives Modell für beispielhafte virale Proteinkonstrukte aufgestellt und die Frage untersucht, welchen Anteil native Proteine und DRiPS zur Antigenpräsentation beitragen. Es wird weiterhin untersucht, inwieweit DRiP Proteine die Antigenpräsentation verändern.

Es werden in Kapitel 3 statistische Modelle für die einzelnen Schritte des MHC I Pathways entwickelt – der Transport von Peptide in das ER (Kapitel 3.2), die Schnittrate des Proteasoms (Kapitel 3.3) und der Einfluss der Aminopeptidasen (Kapitel 3.4). Für die Bindung von Peptiden an MHC I Moleküle existieren bereits gute Modelle (Kapitel 3.1), welche in dieser Arbeit verwendet werden. In Kapitel 3.5 werden die verschiedenen Modelle des Pathways miteinander kombiniert und die Ergebnisse der Antigenvorhersage analysiert.

Die proteasomale Generierung von Peptidfragmenten wird in Kapitel 4 kinetisch modelliert. Die Kombination der erstellten statistischen und kinetischen Modelle ermöglicht die kinetische Simulation der Prozesse zur Präsentation von Antigenen an der Zelloberfläche aus cytosolischen Proteinen (Kapitel 5).

In dieser Arbeit wird vorrangig Wert darauf gelegt, dass die Beiträge der einzelnen Teile des Pathways zur Antigenpräsentation verstanden werden. Daher werden die Modelle mit *in vitro* Daten der isolierten Prozesse trainiert, solche Modelle werden hier als *in vitro* Modelle bezeichnet. Demgegenüber ist es nicht möglich mit Modellen, die auf *in vivo* Daten des gesamten Präsentationsweges basieren, d. h. lediglich der Liste der bekannten präsentierten Epitope sowie der ursprünglichen Proteine, zwischen den Beiträgen der einzelnen Elemente des Pathways zu diskriminieren. Derartige Modelle haben eine relativ hohe Qualität bei der Vorhersage von Antigenen aus der Aminosäuresequenz von Proteinen und werden in dieser Arbeit zum Vergleich der Güte der Vorhersagen herangezogen. In dieser Arbeit werden Modelle, die nur auf *in vivo* Daten beruhen als *in vivo* Modelle bezeichnet.

Das Erreichen einer hohen Vorhersagequalität der Modelle ist in dieser Arbeit zunächst zweitrangig, jedoch ist die Güte der Vorhersagen ein gutes Maß, mit dem man zwischen der Qualität von Modellen diskriminieren kann.

2 Der Beitrag von DRiPs zur Antigenpräsentation

In diesem Kapitel soll die Frage erörtert werden, welchen Anteil so genannte DRiPs an der Proteinsynthese sowie der Antigenpräsentation haben. Es wird ein stark vereinfachtes Modell zum zellulären Proteinumsatz vorgestellt, das die zellulären Prozesse zwischen viraler Infektion und Antigenpräsentation abbildet.

2.1 Hintergrund und experimentelle Daten

2.1.1 Experimenteller Aufbau

Um den Beitrag von DRiPs zur Antigenpräsentation zu quantifizieren hat die Gruppe von Yewdell aufwendige, zeitlich aufgelöste quantitative Messungen von Proteinmengen in Zellen und gleichzeitig Antigenen an der Zelloberfläche durchgeführt[30]. Hierzu wurden von den Experimentatoren verschiedene Proteinkonstrukte erzeugt, die dann mittels eines rekombinanten Vaccinia Virus in Maus L929 Zellen eingebracht wurden. Die Zellen wurden mit cDNA für das MHC-I Molekül H2-K^b transfiziert, so dass die von diesen Molekülen präsentierten Antigene, insbesondere SIINFEKL, gemessen werden konnten. Für die Modellierung waren die folgenden drei Proteinkonstrukte wichtig. Zum ersten wurde ein Kernprotein aus dem Influenza Virus (NP) über einen Linker mit dem aus Ovalbumin stammenden Peptid SIINFEKL verbunden und C-Terminal mit einem grün fluoreszierendem Protein (GFP) verbunden (NP-SIINFEKL-GFP). Beim zweiten Konstrukt wurde zusätzlich in das NP ein Peptid an Stelle 333 inseriert (KEKE⁴-NP-SIINFEKL-GFP), was zu einer veränderten Faltung und damit einhergehenden kürzeren Lebenszeit des Proteins führt[71]. Für das dritte Protein wurde das initiale Methionin des Influenza Kernproteins durch Arginin ersetzt, das N-Terminal mit Ubiquitin verbunden ist. Das Ubiquitin wird noch während der Translation des restlichen Proteinkonstrukts abgespalten, wodurch ein Protein mit sehr kurzer Lebenszeit entsteht (R-NP-SIINFEKL-GFP). In Tabelle 1 sind die chimären Proteine nochmals zusammengefasst. Das beim proteasomalen Abbau der Proteine freigesetzte Peptid SIINFEKL wird präsentiert und ermöglicht das quantifizieren der so erzeugten Antigene an der Zelloberfläche mittels spezifischer Antikörper. Das GFP ermöglicht das Bestimmen der zellulären Proteinkonzentration mittels photometrischer Methoden, wobei jedoch zu beachten ist, dass das GFP nach der Synthese und Faltung erst noch einer autokatalytischen Aktivierung bedarf, bevor es fluoresziert[72, 73, 74]. Mittels quantitativem Western Blot wurde die gemessene Fluoreszenz mit der vorhandenen Pro-

⁴ Das Peptid ist ein 29 Aminosäuren umfassendes Oligomer, das zum großen Teil aus Lysin und Glutamat besteht (Aminosäuresequenz: KEKEKNKLKRKKLENKDKKDEERNKIREE).

teinmenge korreliert und mittels radioaktiver ^{35}S -Methionin Markierung und Messung bestätigt.

Tabelle 1 Übersicht über die chimären Proteinkonstrukte, für die Modellsimulationen ausgeführt und mit den kinetischen Daten aus [30] verglichen wurden

Bezeichnung	Zusammensetzung	berichtete Halbwertszeit
NP-SIINFEKL-GFP	NH_2 -NP-SIINFEKL-GFP-COOH	mehrere Stunden
KEKE-NP-SIINFEKL-GFP	NH_2 -KEKE-NP-SIINFEKL-GFP-COOH	≈ 10 min
R-NP-SIINFEKL-GFP	NH_2 -Ub-R-NP-SIINFEKL-GFP-COOH	≈ 70 min

NP: Influenza Nukleoprotein; GFP: Green Fluorescent Protein; Ub: Ubiquitin, KEKE: Abkürzung für das Insert des Polypeptids in das NP; SIINFEKL: Aminosäuresequenz eines aus dem Ovalbumin stammenden Antigens; NH_2 : N-Terminus des Konstrukts; COOH: C-Terminus des Konstrukts

2.1.2 Experimentelle Ergebnisse

Proteinmengen und Antigene an der Zelloberfläche wurden für die drei Proteinkonstrukte alle 10 Minuten über einen Zeitraum von 5 Stunden gemessen. Zusätzlich wurden Messungen durchgeführt, bei denen die Proteinsynthese nach 140 Minuten mit Cycloheximid inhibiert wurde.

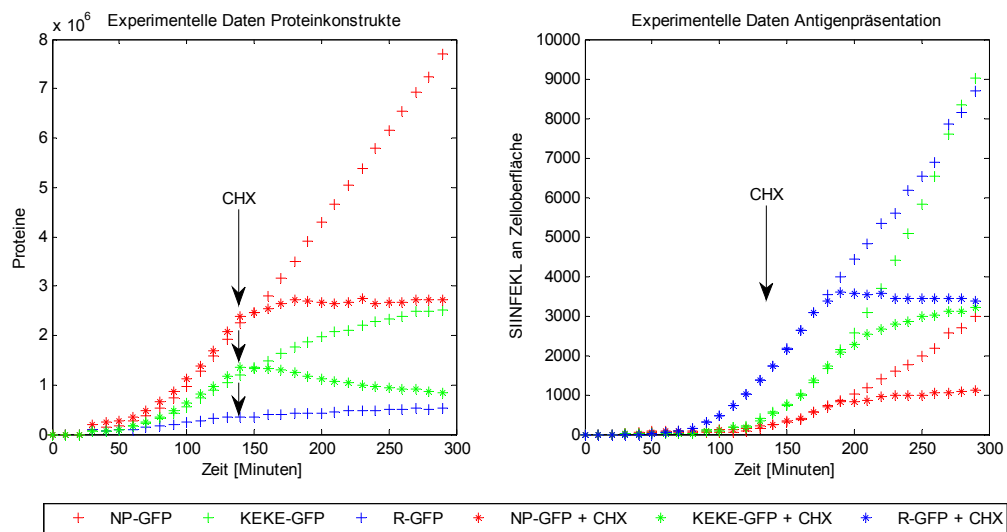


Abbildung 5: Experimentelle Messung der Proteinmengen aus[30]

Links ist der Zeitverlauf der Proteinmengen der drei chimären Proteine, rechts die Menge der gemessenen SIINFEKL Antigenkomplexe an der Zelloberfläche dargestellt. Rot ist in beiden Abbildungen NP-SIINFEKL-GFP, KEKE-NP-SIINFEKL-GFP grün und R-NP-SIINFEKL-GFP blau. Die Messreihe mit Zugabe von Cycloheximid (CHX) nach 140 Minuten ist mit Sternen (*) wiedergegeben, die Messreihe ohne Inhibition der Proteinsynthese mit Kreuzen (+). Für R-NP-SIINFEKL-GFP mit Hemmung der Proteinsynthese gibt es keine experimentellen Daten für die Proteinkonstrukte.

Aus ihren Experimenten folgern die Autoren eine DRiP Rate durchschnittlich $27.1\% \pm 5.9\%$. Diese Schlussfolgerung soll durch kinetische Modellierung in den folgenden Abschnitten genauer untersucht werden.

2.2 Modell und Methoden

2.2.1 Das kinetische Modell

Das hier erstellte kinetische Modell beschreibt die Vorgänge, die direkt nach der Transfektion der Zellen durch das Vaccina Virus auftreten. Es enthält den Prozess der Transkription, wodurch das eingebrachte genetische Material in mRNA Moleküle umgeschrieben wird. Die mRNA wird durch die Ribosomen in Proteine translatiert, wobei eine der beiden neu synthetisierten Fraktionen fehlerhaft ist (DRiP) und somit auch keine Fluoreszenz zeigt, während die GFP Untereinheit der nicht defekten Proteine eine Aktivierung untergeht und dadurch fluoreszierend wird. Alle Proteinfraktionen werden durch das Proteasom abgebaut und ein Teil der erzeugten Fragmente wird an der Zelloberfläche präsentiert.

Das Erzeugen der mRNA wird als Prozess nullter Ordnung, das heißt mit konstanter Rate, modelliert. Der Abbau der mRNA sowie der verschiedenen Proteinfraktionen wird als Prozess erster Ordnung dargestellt. Die Präsentation der SIINFEKL Fragmente an der Zelloberfläche wird ebenfalls als linearer Prozess modelliert, wobei hier jedoch eine Verzögerung zwischen Abbau der Proteine und Erscheinen der Antigene an der Zelloberfläche in das Modell implementiert ist. Das Verschwinden der Antigene von der Zelloberfläche wird nicht modelliert, da dieser Vorgang im experimentellen Zeitfenster nicht mit erfasst wurde. Alle im Modell enthaltenen Vorgänge sind schematisch in Abbildung 6 dargestellt.

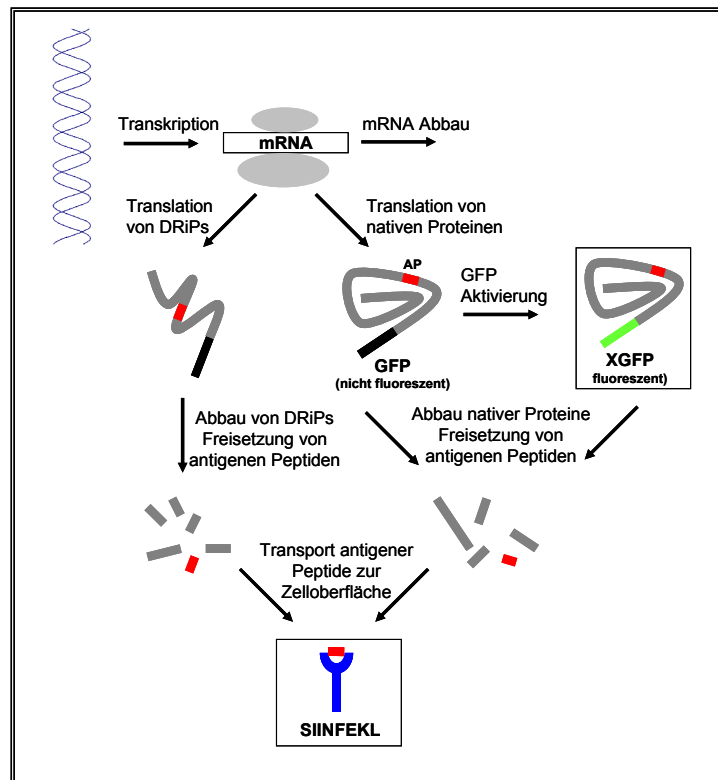


Abbildung 6: Vereinfachtes Reaktionsschema für den zellulären Proteinumsatz und die MHC I vermittelte Antigenpräsentation.

Modellierte Prozesse des Proteinumsatzes nach Transfektion und die MHC I vermittelte Antigenpräsentation. Variablen: mRNA - Messenger RNA des spezifischen in die Zelle transfizierten Proteins; DRiP - von der mRNA translatierte, defekte ribosomale Produkte; GFP - noch nicht fluoreszente, mit GFP markierte, native Protein; XGFP - fluoreszente Form des GFP markierten, nativen Proteins; SIINFEKL - aus den transfizierten Proteinen gebildeten, präsentierten Antigene. Proteolytische Prozesse, die zu einem Verlust von Epitopen oder Vorläufern von Epitopen führen sind im Schema nicht dargestellt, aber im Modell durch den Parameter Präsentationseffizienz η repräsentiert.

Folgende Differentialgleichungen wurden für das Modell benutzt:

(1)

$$\begin{aligned}
\frac{d}{dt}[mRNA] &= v_0 - b_0 * [mRNA] \\
\frac{d}{dt}[DRiP] &= \Theta(T_c - t) * k_{21} * [mRNA] - a_1 * [DRiP] \\
\frac{d}{dt}[GFP] &= \Theta(T_c - t) * k_{22} * [mRNA] - (k_3 + a_2) * [GFP] \\
\frac{d}{dt}[XGFP] &= k_3 * [GFP] - a_2 * [XGFP] \\
\frac{d}{dt}[SIINFEKL] &= \eta * a_1 * [DRiP(t - \tau)] + \eta * a_2 * ([GFP(t - \tau)] + [XGFP(t - \tau)]) \\
\Theta(X) &= \begin{cases} 1 & \text{falls } X \geq 0 \\ 0 & \text{falls } X < 0 \end{cases}
\end{aligned}$$

Die Bedeutung der einzelnen Parameter ist in Tabelle 2 dargestellt. Dieses Differentialgleichungssystem kann analytisch gelöst werden. Unter der Voraussetzung, dass die Initialkonzentrationen der Komponenten zu Beginn der Simulation ($t=0$) gleich Null sind ergeben sich die folgenden Systemgleichungen.

Ohne Inhibition der Protein Synthese ($t < T_c$):

(2)

$$\begin{aligned}
mRNA(t) &= \frac{v_0}{b_0} * [1 - \exp(-b_0 * t)] \\
DRiP(t) &= \frac{v_0 * k_{21}}{a_1 - b_0} * \left[\frac{1 - \exp(-b_0 * t)}{b_0} - \frac{1 - \exp(-a_1 * t)}{a_1} \right] \\
GFP(t) &= \frac{v_0 * k_{22}}{a_2 + k_3 - b_0} * \left[\frac{1 - \exp(-b_0 * t)}{b_0} - \frac{1 - \exp(-(a_2 + k_3) * t)}{a_2 + k_3} \right] \\
XGFP(t) &= \frac{v_0 * k_{22} * k_3}{a_2 + k_3 - b_0} * \left[\frac{1 - \exp(-b_0 * t)}{b_0(a_2 - b_0)} + \frac{1 - \exp(-(a_2 + k_3) * t)}{k_3(a_2 + k_3)} \right. \\
&\quad \left. + \frac{(a_2 + k_3 - b_0) * (1 - \exp(-a_2 * t))}{a_2 * k_3 * (b_0 - a_2)} \right]
\end{aligned}$$

(3)

$$\begin{aligned}
SIINFEKL(t) = & v_0 * a_1 * k_{21} * \eta * \left[\frac{t - \tau}{a_1 * b_0} + \frac{1 - \exp(-b_0 * (t - \tau))}{b_0^2 * (b_0 - a_1)} \right] \\
& + v_0 * a_2 * k_{22} * \eta * \left[\frac{t - \tau}{a_2 * b_0} + \frac{1 - \exp(-b_0 * (t - \tau))}{b_0^2 * (b_0 - a_2)} \right] \\
& - \frac{1 - \exp(-a_1 * (t - \tau))}{a_1^2 * (b_0 - a_1)} \\
& - \frac{1 - \exp(-a_2 * (t - \tau))}{a_2^2 * (b_0 - a_2)}
\end{aligned}$$

Und nach Einsatz der Inhibition:

(4)

$$\begin{aligned}
DRiP^{\otimes}(t) &= DRiP(T_c) * \exp(-\gamma * a_1 * (t - T_c)) \\
GFP^{\otimes}(t) &= GFP(T_c) * \exp(-(\gamma * a_2 + k_3) * (t - T_c)) \\
XGFP^{\otimes}(t) &= XGFP(T_c) * \exp(-\gamma * a_2 * (t - T_c)) \\
&+ GFP(T_c) * \exp(-\gamma * a_2 * (t - T_c)) * (1 - \exp(-k_3 * (t - T_c)))
\end{aligned}$$

(5)

$$\begin{aligned}
t \leq T_c + \tau : \\
SIINFEKL^{\otimes}(t) &= SIINFEKL(t) \\
t \geq T_c + \tau : \\
SIINFEKL^{\otimes}(t) &= SIINFEKL(T_c + \tau) \\
&+ \eta * DRiP(T_c) * (1 - \exp(-\gamma * a_1 * (t - T_c - \tau))) \\
&+ \eta * (GFP(T_c) + XGFP(T_c)) * (1 - \exp(-\gamma * a_2 * (t - T_c - \tau)))
\end{aligned}$$

Die Konstante γ berücksichtigt die Änderung der Proteasomaktivität bei Hemmung der Proteinsynthese (siehe Ergebnisteil dieses Kapitels).

Alternativ zur Repräsentation der Prozessierung der proteasomal erzeugten Proteinfragmente im MHC-I Pathway durch das verzögernde Argument τ , gibt es auch eine Variation des Modells, bei der die einzelnen Schritte der Antigenprozessierung durch eine Kette von linearen Differentialgleichungen dargestellt wird. Nach der Produktion der durch das Proteasom durchlaufen sie eine Reihe von Zwischenschritten, bis sie an der Zelloberfläche

präsentiert werden (Protein $\rightarrow X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_n \rightarrow$ präsentiertes Antigen). In dieser Modellvariante wird die Differenzialgleichung für SIINFEKL in Gleichung 1 ersetzt durch:

(6)

$$\begin{aligned}\frac{d}{dt}[I_1] &= \eta * a_1 * [DRiP] + \eta * a_2 * ([GFP] + [XGFP]) - k_{mean} * [I_1] \\ \frac{d}{dt}[I_2] &= k_{mean} * [I_1] - k_{mean} * [I_2] \\ &\dots \\ \frac{d}{dt}[SIINFEKL] &= k_{mean} * [I_n]\end{aligned}$$

Die Präsentationseffizienz, also der Anteil der SIINFEKL-Fragmente, die wirklich als MHC Komplexe an die Zelloberfläche gelangen, ist in die erste Gleichung eingefügt, wobei es aber keine Rolle spielt, ob es verschiedene Faktoren für jeden Schritt der Antigenpräsentation oder, wie hier, nur einen Koeffizienten gibt. Anstatt der konstanten Verzögerung der Antigenpräsentation τ gibt es in dieser Variante die durchschnittliche Rate k_{mean} mit der das zukünftige Antigen die Kette durchläuft sowie die Anzahl der in der Kette enthaltenen Schritte n .

2.2.2 Anpassen der Modellparameter an die experimentellen Daten

Die Werte der Modellparameter wurden bestimmt, indem der Abstand (Gleichung 7) zwischen den experimentellen Werten und den vom Modell erzeugten theoretischen Werte minimiert wurden. Die Abstandsfunktion bestand aus der Summe der quadrierten Differenzen zwischen den experimentellen Werten und den Modellergebnissen zu allen Zeitpunkten, an denen gemessen worden ist. Zusätzlich wurden die Abstände der einzelnen Kurven gewichtet in den Gesamtabstand einbezogen, da sich die Werte der verschiedenen gemessenen Proteinmengen und Epitopzahlen um mehrere Größenordnungen unterscheiden.

(7)

$$Dist = \sum_{Messreihen} \frac{\sum_{Datenpunkte} (Datenpunkt - Simulierter\ Wert)^2}{Maximum(Messreihe)^2}$$

Das Anpassen der Modellparameter an die experimentellen Daten wurde mittels Microsoft Excel durchgeführt (Solver 5.5 für Excel von Frontline Systems).

Da die einzelnen Proteinkonstrukte in ihrer Länge und Sequenz sehr ähnlich sind, sollten auch die Parameter, die die Synthese der mRNA sowie die gesamte Proteinsynthese beschreiben ($v_0 * (k_{21} + k_{22})$), identische Werte annehmen. Die Rate der GFP Aktivierung (k_3)

sowie die Verzögerung der Antigenpräsentation nach Abbau der Proteine (τ) als auch die Abbaugeschwindigkeit der mRNA (b_0) sollten ebenfalls unabhängig vom Proteinkonstrukt sein und wurden ebenfalls für chimären Proteine auf die gleichen Werte fixiert. Die Syntheserate der mRNA tritt in den Modellgleichungen immer im Produkt mit der Syntheserate der Proteine (k_{21} und k_{22}) auf und kann nicht allein bestimmt werden, da keine Messungen der mRNA Mengen vorliegen. Die Effizienz der Antigenpräsentation nach Abbau der Proteine (η) wurde ebenfalls für alle Proteinkonstrukte auf einen gemeinsamen Wert festgesetzt. Die Abbaurate für DRiPs wurde auf eine Halbwertszeit von 15 Minuten festgelegt ($a_1 = 4.62 \times 10^{-2}$ min).

Durch diese Reduktion der Anzahl der frei bestimmbaren Parameter wird „Overfitting“ – das Überanpassen der Modellparameter an die experimentellen Daten – vermindert und die Werte der einzelnen Parameter können besser interpretiert werden, da zusätzliche Freiheitsgrade dazu führen können, dass Änderungen an einem Parameter durch Änderungen der anderen Parameter kompensiert werden und somit zwei verschiedene Parametersätze das gleiche Verhalten des Systems beschreiben.

2.3 Ergebnisse

2.3.1 Nur einige experimentelle Kurven lassen sich durch das Modell reproduzieren.

Mit dem im Kapitel 2.2.1 beschriebenen Modell wurde eine Reihe von Rechnungen durchgeführt, wobei die Konstante γ nicht implementiert war beziehungsweise den festen Wert 1.0 hatte. Hierbei wurde versucht, die experimentellen Zeitverläufe der Proteinmengen und präsentierten Antigene möglichst gut mit den vom Modell erzeugten Kurven wiederzugeben. Durch Minimierung des Abstandes zwischen Modellergebnissen und experimentellen Messungen wurde ein optimaler Parametersatz bestimmt. Mit diesem können die meisten experimentellen Kurven relativ gut wieder gegeben werden, jedoch zeigen sich bei einzelnen Kurven systematische Abweichungen zwischen den Simulationen und dem Experiment. In Abbildung 7 ist dargestellt, dass für das NP-SIINFEKL-GFP Konstrukt die Proteinkonzentrationen sehr gut reproduziert werden können, jedoch nicht die Menge der präsentierten Epitope ohne Hemmung der Proteinkonzentration. Bei der Abbildung von KEKE-NP-SIINFEKL-GFP ist zu erkennen, dass jeweils nur eine der beiden Kurven für Proteinmengen und Antigenanzahl gut reproduziert werden konnte. Das jeweils die Abweichungen in den Kurven ohne Hemmung der Proteinsynthese zu sehen sind, liegt an der Auswahl des Parametersatzes. Es ist durchaus möglich einen anderen Parametersatz zu finden der einen ähnlich guten Gesamtabstand zwischen Modell und Experiment aufweist,

wie der hier gewählte und die Kurven ohne Hemmung der Proteinsynthese gut reproduziert, dafür jedoch nicht mehr die Kurven mit Hemmung der Proteinsynthese beschreiben kann.

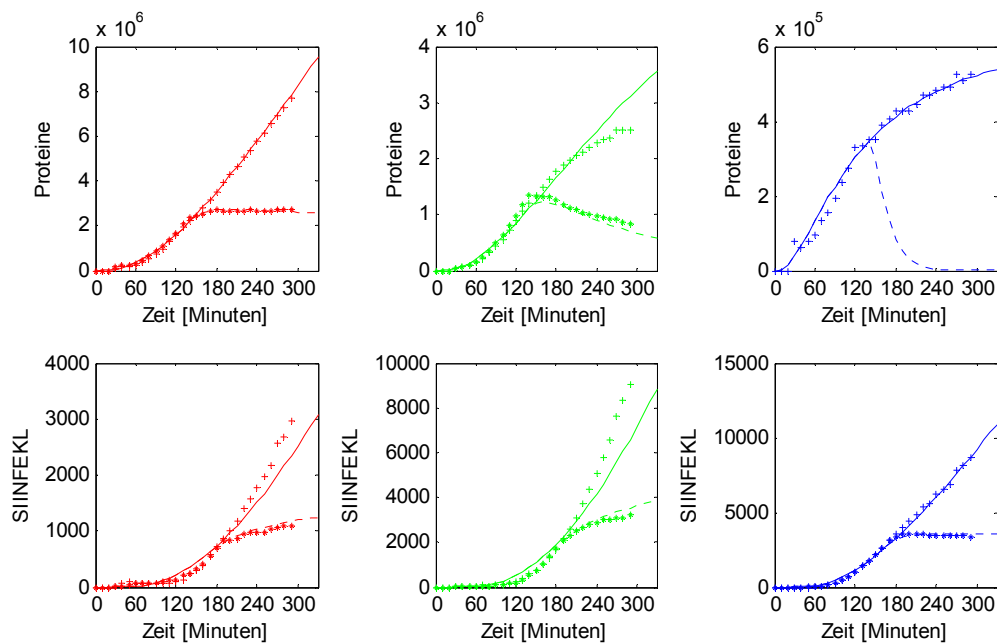


Abbildung 7: Proteinmengen und präsentierte Antigene mit Modellvariante ohne γ

Links: NP-SIINFEKL-GFP; Mitte: KEKE-NP-SIINFEKL-GFP; rechts: R-NP-SIINFEKL. Die oberen Paneele zeigen jeweils den zeitlichen Verlauf der Proteinmengen und die unteren die Menge der an der Zelloberfläche präsentierten Epitope. Die experimentellen Daten sind wie in Abbildung 5 aufgetragen. Die Simulationsergebnisse sind durch Linien dargestellt, wobei die gestrichelte Linie die Simulation mit Hemmung der Proteinsynthese und die durchgehende Linie ohne Inhibition der Proteinsynthese ist.

Wenn man die Einfachheit des Modells in betracht zieht, sind diese Abweichungen nicht übermäßig bedenklich. Jedoch weisen diese systematischen Differenzen auf Eigenschaften des experimentellen Systems hin, die das Modell in dieser Form nicht reproduzieren kann.

2.3.2 Verschiedene Modellvariationen wurden getestet.

Zur Erklärung der verbliebenen Unterschiede zwischen den Modellsimulationen und den experimentellen Ergebnissen wurden verschiedene Hypothesen aufgestellt, dann in die Gleichungen des Modells integriert und so getestet, ob diese Hypothese in der Lage ist, die Abweichung zu erklären.

Hypothese 1: Bei der Transfektion der Zellen tritt eine Verzögerung auf bis es zur Synthese von mRNA kommt.

Nachdem die Mauszellen den Viren ausgesetzt werden, müssen diese in die Zellen eindringen und ihr Genom der Transkriptionsmaschinerie der Wirtszelle zur Verfügung stellen, um mRNA für virale Proteine herstellen zu lassen. Ein zusätzlicher Parameter wurde ähnlich zum schon implementierten τ als verzögerndes Element implementiert. Mit dieser Veränderung des Modells konnte keine wesentliche Verbesserung der systematischen Abweichung erzielt werden. Somit zeigt sich, dass der Prozess der Transfektion durch das Modell ausreichend gut abgebildet wird, zumal keine Messungen der mRNA Mengen vorliegen.

Hypothese 2: Die Präsentation der Epitope an der Zelloberfläche ist im Modell zu einfach implementiert.

Zwischen dem Abbau der Proteine durch das Proteasom und der Präsentation der Antigene an der Zelloberfläche liegt ein vielschrittiger Prozess, der unter anderem aus dem Transport der Peptide in das ER mittels TAP, dem Binden der Peptide an die MHC-Moleküle, dem Trimmen der Peptide durch Aminopeptidasen sowie dem Transport der MHC-Molekül-Antigen-Komplexe zur Zelloberfläche besteht[75]. Allein der letzte Schritt läuft über vesikulären Transport vom ER über Cis- und Transgolgi zur Zelloberfläche. Das Auffüllen dieser intermediären Epitoppools könnte für die beobachtete Verzögerung der Präsentation der Antigene nach dem Abbau der Proteine verantwortlich sein. In der Annahme, dass die einfache Repräsentation dieses Prozesses durch ein verzögerndes Element in Gleichung 1 diese Vorgänge übermäßig vereinfacht, wurde diese Gleichung durch Gleichung 6 ersetzt. Um die Anzahl der Parameter noch in einem handhabbaren Rahmen zu halten gibt es nur die zwei neuen Parameter k_{mean} und n anstelle von τ . Auch diese Variation des Modells führt nicht zu einer qualitativen Verbesserung der Reproduktion der experimentellen Daten durch das Modell. Die im Modell vorgestellte Wiedergabe dieses Prozesses ist zumindest gleichwertig zu der hier getesteten Variante, wobei auch hier anzumerken ist, dass es für die Intermediate im ER und Golgi keine experimentellen Daten gibt.

Hypothese 3: Die Effizienz, mit der Epitope aus DRiPs oder nativen Proteinen gebildet werden, unterscheidet sich voneinander.

Es besteht die Möglichkeit, dass Epitope, die aus DRiPs stammen, mit einer anderen Effizienz prozessiert werden, als Epitope, die aus nativen Proteinen erzeugt werden[76]. Simulationen, die mit dieser Variante des Modells durchgeführt wurden, sind nicht we-

sentlich besser in der Lage die experimentellen Ergebnisse zu reproduzieren, als die Annahme der gleichen Effizienzen für DRiPs und native Proteine. Auf diesen Punkt wird in der Diskussion weiter eingegangen.

Hypothese 4: Bei der Applikation von Cycloheximid stirbt ein Teil der Zellen ab.

Die Applikation von Cycloheximid sowie die damit verbundene Hemmung der Proteinsynthese ist eine massive Störung der Zellen und die Annahme ist hier, dass ein Teil der Zellen abstirbt und somit keine Änderung der Proteinmengen und Epitope mehr durchführt.

Dieses Modell kann die experimentellen Daten deutlich besser reproduzieren als alle bisher betrachteten Varianten (Abbildung 8). Es bleiben noch immer einzelne Abweichungen zu den gemessenen Proteinmengen und Epitopzahlen aber die Ergebnisse sind insgesamt zufrieden stellend. Problematisch ist jedoch, dass hierzu 56% der Zellen durch Cycloheximid zerstört werden, was von den Experimentatoren nicht festgestellt wurde. Auch ist die Annahme des sofortigen kompletten Einstellens der Tätigkeit der Zellen, anstatt der allmählichen Abnahme der Vitalität, schwer zu rechtfertigen.

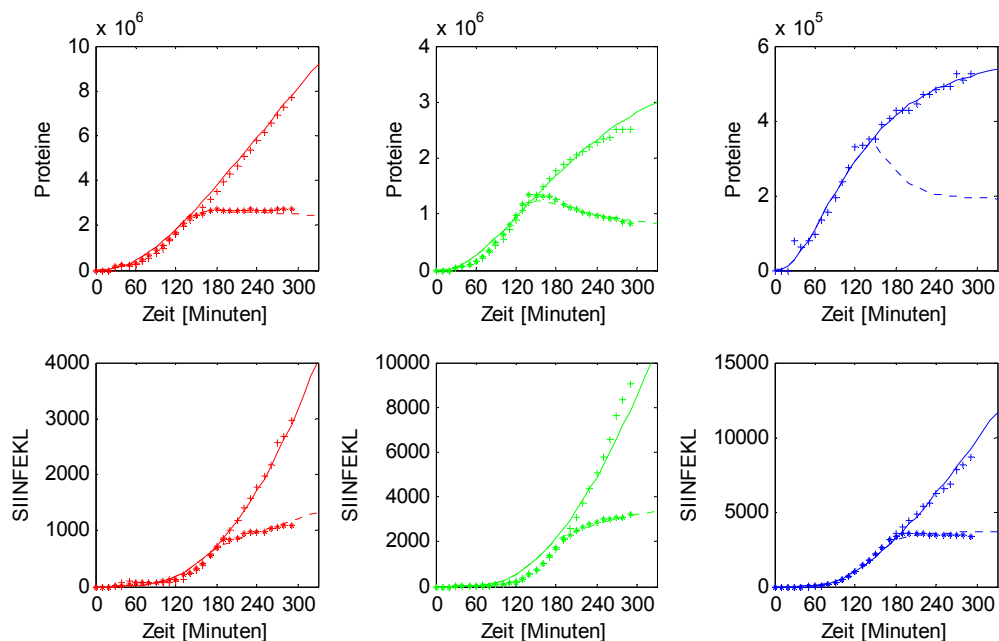


Abbildung 8: Proteinmengen und präsentierte Antigene bei Annahme sofortiger massiver Zellschädigung durch Cycloheximid

Auftragung ist wie in Abbildung 7. Für eine Fraktion der Zellen wird nach Applikation von Cycloheximid angenommen, dass die Proteinmengen und Antigene konstant bleiben.

Alle hier beschriebenen Rechnungen sind somit nicht in der Lage, alle Experimente gleichzeitig wiederzugeben. Man kann aber als Ergebnis ansehen, dass sie aufzeigen, welche Bereiche des Modells ausreichend gut beschrieben sind und wo Änderungen ansetzen müssen, um erfolgreiche Resultate zu erzielen. So zeigt sich, dass die Produktion der mRNA hinlänglich gut die Realität abbildet. Auch die Prozesse des Transports der Antigene zur Zelloberfläche als auch die Effizienzen der Produktion der Antigene werden durch hinzufügen von Parametern und damit Freiheitsgraden nicht besser beschrieben beziehungsweise ermöglichen keine wesentlich bessere Anpassung an die Daten. Erst eine Änderung die sowohl den Abbau der Proteine als auch die Generierung von Antigenen beeinflusst, bewirkt hier eine erhebliche Veränderung. Hervorzuheben ist auch das die Hemmung der Proteinsynthese über den primären Effekt hinaus das Verhalten des Systems ändert.

2.3.3 Hemmung der Proteinsynthese verringert die proteasomale Abbaurate

Die zuvor beschriebenen Überlegungen führten zu der Annahme, dass bei der Inhibition der Proteinsynthese durch Cycloheximid ebenfalls der proteasomale Proteinabbau reduziert wird.

Hypothese 5: Inhibition der Proteinsynthese führt zu einer Reduktion des Proteinabbaus.

Im Modell beschreibt der Parameter γ die Änderung der Abbauraten der verschiedenen Proteinpools. Da die Annahme ist, dass die Aktivität des Proteasoms generell reduziert wird, wirkt der Reduktionsparameter γ in gleicher Weise auf die Raten a_1 und a_2 . Eine weitere implizite Annahme ist die sofortige Verringerung der proteasomalen Aktivität nach Applikation von Cycloheximid beziehungsweise sollte die Übergangsdauer zwischen den beiden Zuständen unterhalb der durch das Experiment festgelegten zeitlichen Auflösung von 10 Minuten liegen.

Der optimale Wert von $\gamma = 0.35$, also die Reduktion der proteasomalen Aktivität auf etwa 1/3 des normalen Wertes, beseitigt die meisten der systematischen Abweichungen zwischen den experimentellen Werten und der Modellsimulation. Der Abstand zwischen den beiden Progresskurven wurde durch das Einführen dieses Parameters etwa halbiert. Dieses Ergebnis ist in Abbildung 9 dargestellt

Eine anschließende Literatursuche nach einer solchen Wirkung von Cycloheximid lieferte eine Reihe von Arbeiten, die die Hypothese unterstützen[77, 78, 79, 80, 81, 82, 83].

Hierbei ist jedoch festzuhalten, dass Cycloheximid nicht direkt auf das Proteasom wirken muss. Wahrscheinlicher ist, dass das zelluläre System der Proteinsynthese in ständiger Kommunikation mit dem Abbau steht und somit die Aktivität des proteasomalen Systems kontrolliert, jedoch ist der Vermittler dieser Information unbekannt.

2.3.4 Auswertung der Modellparameter

In Tabelle 2 sind die optimalen Parameter zum Beschreiben der experimentellen Daten durch das Modell dargestellt. Aus den Parameterwerten können eine Reihe von Schlussfolgerungen gezogen werden:

- Die maximale Proteinsyntheserate $v_0/b_0 \times (k_{21} + k_{22})$ beträgt 9.40×10^4 Proteine pro Minute, wobei aber innerhalb der 5 Stunden des Experiments nur 78% dieser Rate erreicht werden. Der Bereich entspricht auch der von Yewdell bestimmten Proteinsyntheserate von 36'000 Proteinen pro Minute (persönliche Kommunikation), welche in der Simulation nach etwa 90 Minuten erreicht wird.
- Die anfängliche Änderung der Proteinsyntheserate beträgt $v_0 \times (k_{21} + k_{22}) = 472$ Proteine pro Minute² und geht mit der Halbwertszeit $\ln(2)/b_0 = 138$ Minuten gegen Null. Um abzuschätzen, inwieweit dieser Wert realistisch ist, wurden folgende Überlegung angestellt: Die Länge der Proteinkonstrukte beträgt etwa 700 Aminosäuren und somit besteht das entsprechende mRNA Transkript aus etwa 2100 Nukleotiden. Bei einer durchschnittlichen Translationsrate von 7 Aminosäuren pro Sekunde und einer Transkriptionsrate von 70 Nukleotiden pro Sekunde[84] erhält man eine Änderung der Proteinsyntheserate von 0.83 pro Minute² bei Annahme von einem Ribosom pro mRNA Transkript, einem Gen pro Zelle und keinem simultanen Transkribieren von diesem Gen. Die beobachtete Rate liegt um den Faktor 566 höher, was auf einen sehr effizienten Promotor, multipler Translation einer mRNA durch Polysomen[85] und möglicherweise die Existenz mehrerer Gene pro Zelle hin.
- Die Abbaurate der mRNA $b_0 = 5.02 \times 10^{-3}$ entspricht einer Halbwertszeit von 138 Minuten. Dieser Wert zeigt eine hohe Stabilität und liegt im Bereich der berichteten Lebenszeiten von mRNA Transkripten, welche von einigen Minuten bis zu mehreren Stunden variieren[86].
- Die Zeitverzögerung τ zwischen dem Abbau der Proteine und der damit verbundenen Generierung von Antigenen oder N-Terminal verlängerten Vorläufern[87] und dem Erscheinen der beladenen MHC Komplexe an der Zelloberfläche beträgt $\tau = 40.5$ Minuten. Diese lange Verzögerung muss aufgrund von Prozessen im ER und entlang des vesikulären Pathways zustande kommen, da die Lebenszeit von kurzen Peptiden im Cytosol nur wenige Sekunden betragen[88]. Wenn die hier berechnete Verzögerung in

der Größenordnung allgemeingültig für die Antigenpräsentation ist, so ist das Auftauchen von viralen Antigenen an der Oberfläche infizierter Zellen in anderen *in vivo* Systemen nicht vor einer halben Stunde nach Infektion zu erwarten. Dieses ist in guter Übereinstimmung mit der Beobachtung, dass RMA Zellen (Maus Tumor Zellen) von für das nucleocapsid Protein spezifischen CD8 T Zellen lysiert wurden, nachdem sie für 45 Minuten diese Protein produzierten[89].

- Die Halbwertszeit der Aktivierung von GFP beträgt entsprechend der Simulation $\ln(2)/k_3 = 12.2$ Minuten. Dieser Wert ist deutlich kürzer als die von Princiotta et al.[30] angegebene Aktivierungszeit, die direkt aus der beobachteten zeitlichen Differenz der Maxima der Proteinsynthese Messung mittels Fluoreszenz und radioaktiver Markierung abgeleitet wurde. Der Unterschied dieser Werte ist methodischer Natur und demonstriert den Vorteil eines quantitativen Modells. Mit dem Wert des Parameters k_3 in Tabelle 2 beträgt die zeitliche Verschiebung der Maxima von GFP sowie XGFP tatsächlich 30 Minuten. Aus kinetischen Gründen ist diese Verschiebung aber nicht identisch mit der Halbwertszeit der Fluoreszenzaktivierung.
- Die Antigenpräsentationseffizienz wurde berechnet mit $\eta = 1.08 \cdot 10^{-3}$, dass heißt $1/\eta = 926$ Proteine müssen abgebaut werden, um im Durchschnitt ein Antigen an der Zelloberfläche zu produzieren. Dieser Wert liegt nahe an dem von Princiotta et al.[30] angegebenen Wert von 994 Proteinen pro Antigen und liegt ziemlich genau beim geometrischen Mittel der berichteten minimalen und maximalen Effizienz von 100 beziehungsweise 10000 Proteinen pro präsentiertem Antigen[90, 91].
- Wie zuvor ausführlich beschrieben, verbesserte sich die Reproduktion der experimentellen Daten durch das Modell sehr durch die Einführung des Parameters γ , der die proteasomalen Abbauraten a_1 und a_2 , unabhängig vom betrachteten Proteinkonstrukt, nach der Applikation des Proteinsyntheseinhibitors Cycloheximid modifiziert. Der berechnete Wert von $\gamma = 0.35$ bedeutet, dass nach Inhibierung der Proteinsynthese die Proteinabbaukapazität der Zellen auf 35% des Originalwertes sinkt.
- Die Halbwertszeiten der Proteinkonstrukte NP-SIINFEKL-GFP, KEKE-NP-SIINFEKL-GFP sowie R-NP-SIINFEKL-GFP betragen in der Modellsimulation $\ln(2)/a_2 = 249, 44$ und 12 Minuten, was gut mit den zuvor berichteten Werten in Tabelle 1 übereinstimmt.
- Das Verhältnis der Ratenkonstanten für die Synthese der nativen und der defekten Proteine k_{22} und k_{21} gibt den Anteil an, den DRiPs an der gesamten Proteinsynthese haben. Diese DRiP Rate beträgt in der Simulation etwa $k_{21}/k_{22} = 10\%$ für alle drei betrachteten Proteinkonstrukte. Dieser Wert liegt deutlich unter den in der Literatur berichteten Werten von mindestens 35%[30, 31], jedoch wurden diese Werte ohne Zuhilfenahme eines kinetischen Modells abgeschätzt.

Tabelle 2: Parameter des kinetischen Modells für die Simulation der Proteinlevel und präsentierten Antigenen

Parameter	Einheit	Bedeutung	Numerischer Wert		
			NP	KEKE	R
v_0	mol min^{-1}	mRNA Syntheserate *		1	
b_0	min^{-1}	mRNA Abbaurate		5.02×10^{-3}	
k_3	min^{-1}	GFP Fluoreszenzaktivierungsrate		5.69×10^{-2}	
$(k_{21}+k_{22})$	min^{-1}	Gesamtproteinsyntheserate *		4.72×10^2	
η		Präsentationseffizienz		1.08×10^{-3}	
a_1	min^{-1}	Abbaurate DRiPs		4.62×10^{-2}	
γ		Reduktionsfaktor für Proteinabbau		3.50×10^{-1}	
τ	min	Verzögerungszeit zwischen Proteinabbau und Antigenpräsentation		4.05×10^1	
a_2	min^{-1}	Abbaurate natives Protein	2.78×10^{-3}	1.58×10^{-2}	5.72×10^{-2}
k_{21}	min^{-1}	Syntheserate DRiP *	4.43×10^1	5.75×10^1	4.03×10^1
k_{22}	min^{-1}	Syntheserate natives Protein *	4.28×10^2	4.15×10^2	4.32×10^2

Die Syntheseraten von mRNA und Proteinen erscheinen bei der Simulation der beobachtbaren Größen immer als Produkt. Der Wert für die mRNA Syntheserate wurde willkürlich auf 1 gesetzt. Änderungen an v_0 führen zu entsprechenden Änderungen bei k_{21} und k_{22} .

2.3.5 Der Einfluss von DRiPs auf die Antigenpräsentation

Um den tatsächlichen Beitrag von DRiPs für die Antigenpräsentation genauer zu untersuchen, wurden die Zeitverläufe, dargestellt in Abbildung 9 B,D und F, in die Beiträge aus den verschiedenen Proteinpools der nativen und der defekten Proteine aufgespalten und in Abbildung 10 dargestellt. Der Anteil der Antigene, die aus DRiPs stammen, ist nur für das langlebige Konstrukt NP-SIINFEKL-GFP vorherrschend und auch hier nur die relativ frühe Phase zwischen 50 und 100 Minuten nach Transfektion der Zellen. Für die beiden anderen weniger stabilen Konstrukte ist der Beitrag aus dem nativen Proteinpool sogar für diese frühe Phase dominierend. Die Verkürzung der Zeitspanne zwischen Transfektion und dem Erscheinen von für eine Immunantwort ausreichend großen Mengen von Antigenen an der Zelloberfläche ist in Abbildung 11 für die verschiedenen Konstrukte dargestellt. Es ist klar zu erkennen, dass die Beschleunigung der Immunantwort von der Menge der benötigten Antigene abhängt und für kurzlebige Proteine in der Regel deutlich unter 5 Minuten liegt. Selbst bei Proteinen mit Lebenszeiten im Bereich einer Stunde ma-

ximal wird die Antigenpräsentation nur um etwa 5 Minuten beschleunigt. Erst bei stabilen Proteinen wie dem NP-SIINFEKL-GFP ist dieser Prozess wesentlich schneller.

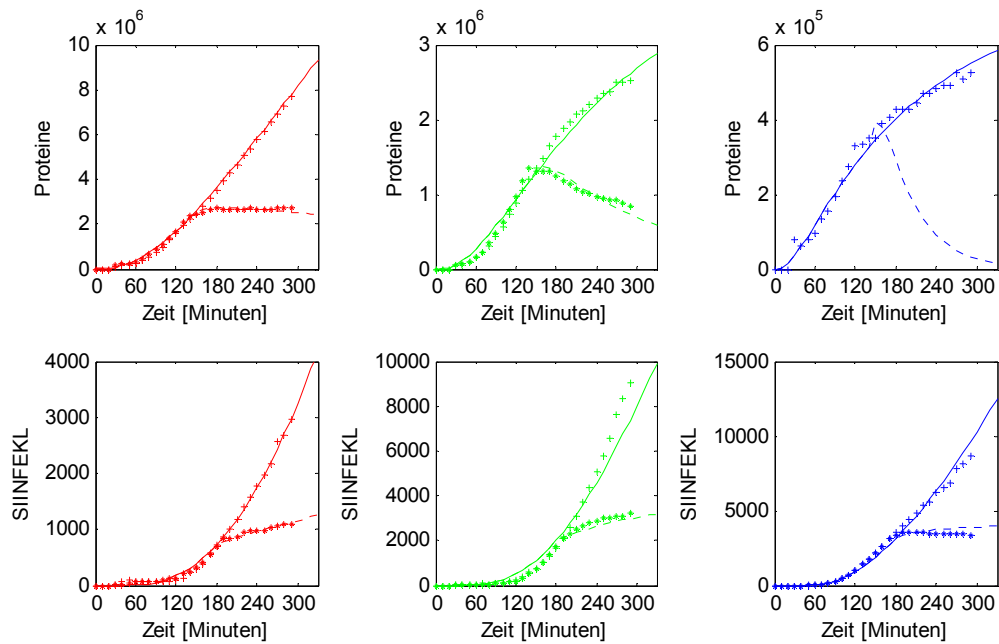


Abbildung 9: Proteinmengen und präsentierte Antigene bei partieller Hemmung des Proteasoms durch Cycloheximid

Auftragung ist wie in Abbildung 7. Die Abbaurrate aller Proteine wird nach Applikation von Cycloheximid um den Faktor $\gamma=0.35$ reduziert. Die verwendeten Parameter sind in Tabelle 1 wiedergegeben.

Ausgehend von den Parameterwerten in Tabelle 2 wurden Zeitverläufe für die Antigenpräsentation von Peptiden aus hypothetischen Proteinen mit unterschiedlichen Lebenszeiten simuliert (Abbildung 12). Für ein sehr stabiles Protein mit einer Halbwertszeit von 10'000 Minuten (etwa 7 Tage) werden während der ersten 5 Stunden nach Transfektion fast alle Antigene aus dem DRiP Pool generiert. Dieser Anteil verringert sich mit sinkender Halbwertszeit der Proteine. Voraussetzung hierfür ist natürlich, dass die an das Experiment von Princiotta et al.[30] Parameter eine typische *in vivo* Situation beschreiben und die Annahme der gleichen Effizienzen der Prozessierung der Antigene für DRiPs und native Proteine identisch ist.

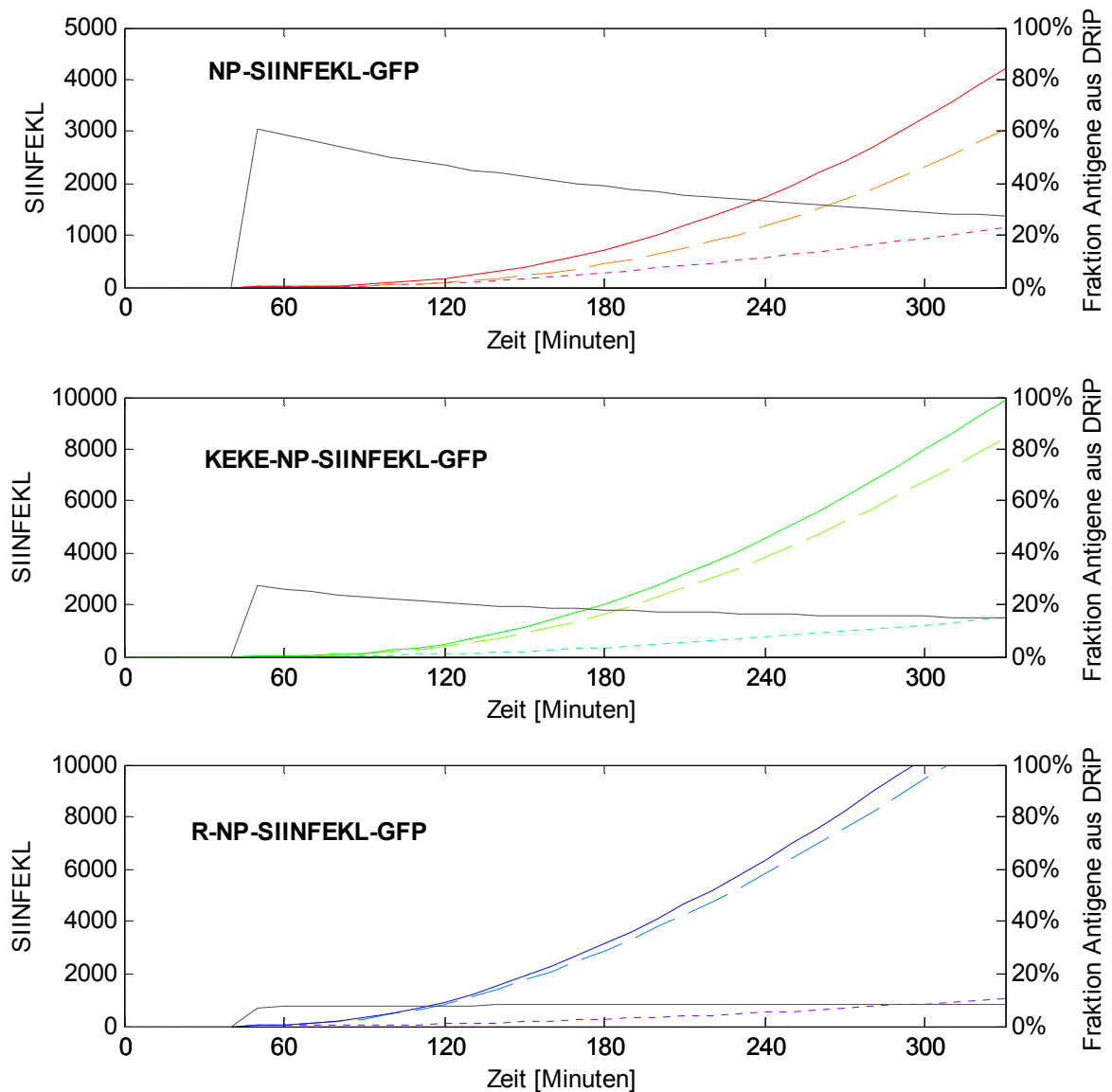


Abbildung 10: Beitrag von DRiPs und nativen Proteinen zur Antigenpräsentation - Protein-konstrukte

Jedes Paneel stellt die Beiträge von Epitopen aus DRiPs (gepunktete Linie) und aus nativen Proteinen (gestrichelt) zur Antigenpräsentation (durchgehend, farbig) dar. Der Prozentsatz der Antigene aus DRiPs ist als graue Linie dargestellt.

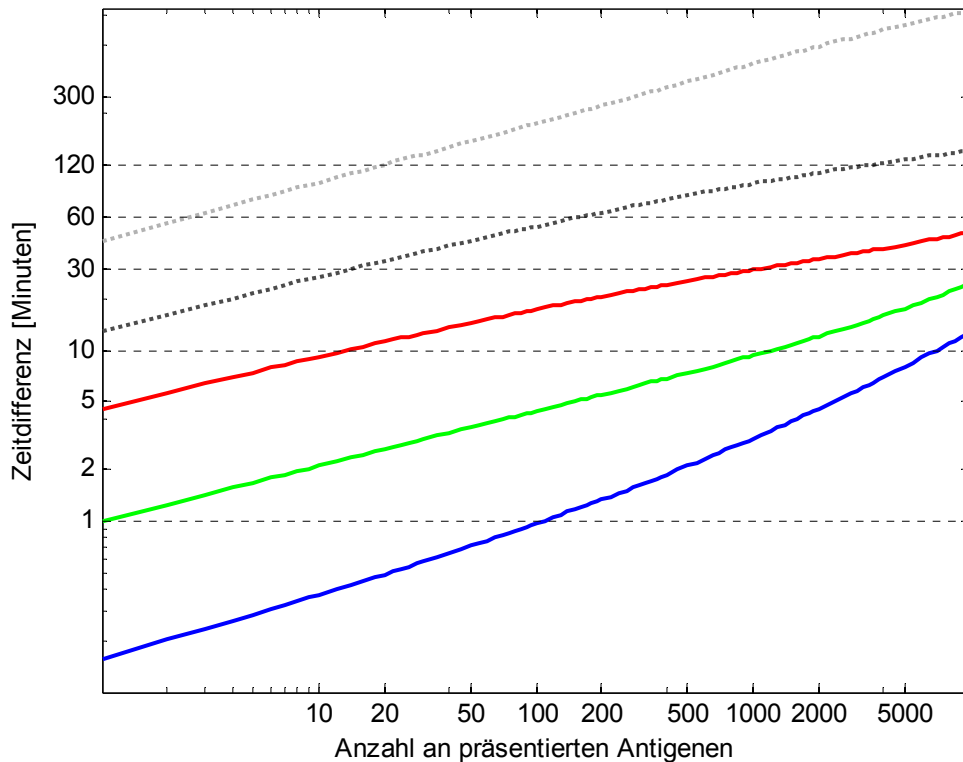


Abbildung 11: Beschleunigung der Immunantwort durch DRiPs

Die Zeitdifferenz für die Präsentation einer gegebenen Anzahl an Antigenen an der Zelloberfläche (x-Achse) mit und ohne Epitope aus DRiPs für die drei chimären Proteine (farbig) und für zwei hypothetische Proteine mit Lebenszeiten von 1'000 (dunkelgrau, gepunktet) und 10'000 Minuten (hellgrau, gepunktet). Beide Achsen überspannen mehrere Größenordnungen und wurden daher logarithmiert. Für die Simulation wurden die Parameter aus Tabelle 2 verwendet.

2.4 Diskussion

Das Ziel der in diesem Kapitel beschriebenen Studie war die Quantifizierung der an der Antigenpräsentation aus endogenen Proteinen beteiligten Hauptprozesse. Die Untersuchung des Einflusses der DRiPs auf die Antigenpräsentation, sowohl qualitativ als auch quantitativ, war dabei die hauptsächliche Fragestellung. Diese Frage sollte durch Untersuchung bereits publizierter, quantitativer experimenteller Daten, insbesondere solcher, die die DRiP Hypothese begründen, durchgeführt werden. Eine Quantifikation der verschiedenen Prozesse, die dem zellulären Proteinumsatz und der MHC Class I Antigenpräsentation zugrunde liegen, bedarf eines kinetischen Modells. Daher wurde ein mechanistisches mathematisches Modell entwickelt, das trotz seiner Einfachheit und dem Zusammenfassen vieler Elementarprozesse zu einer guten Beschreibung der untersuchten experimen-

tellen Daten führt. Aus den Parametern des Modells konnten Schlussfolgerungen gezogen werden, die es erlauben, die gestellte Frage nach der Rolle der DRiPs zu beantworten.

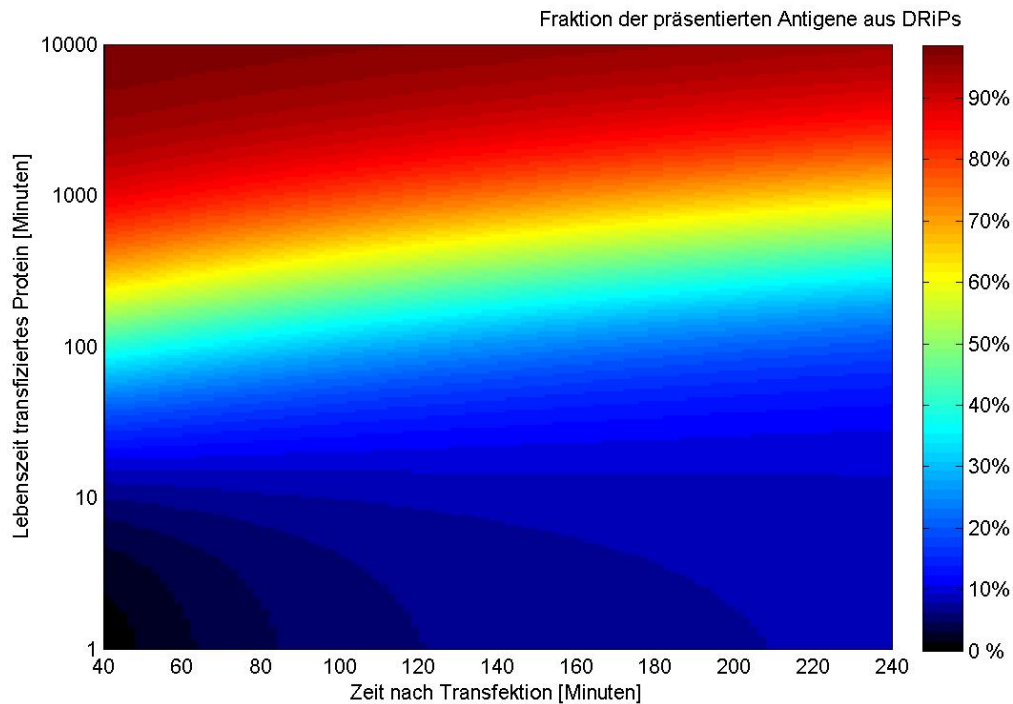


Abbildung 12: Beitrag von DRiPs zur Antigenpräsentation - hypothetische Proteine

Anteil der aus DRiPs generierten Epitope an der gesamten Antigenpräsentation nach Transfektion abhängig von der Lebenszeit der Proteine. Der Anteil ist farbkodiert nach der rechts abgebildeten Skala. Die Werte wurden durch Simulationen mit den Parametern aus Tabelle 2 durchgeführt. Die Lebenszeit der nativen Proteine a_2 wurde systematisch variiert.

2.4.1 Eine DRiP Rate von 10% ist ausreichend für die Antigenpräsentation.

Ein verblüffendes, weil im Gegensatz zur bestehenden Literatur, Ergebnis der Berechnungen ist, das der Anteil der DRiPs an der Proteinsynthese nur etwa 10% in dem untersuchten Transfektionsexperiment beträgt. Jedoch auch diese geringe Rate ist ausreichend um das beobachtete zeitliche Verhalten der präsentierten Antigene zu beschreiben. Der in der Literatur dargestellte Anteil der DRiPs von mehr als 35% wurde aus der Akkumulation von radioaktiv markierten Proteinen nach Zugabe eines Proteasominhibitors gemessen[22]. Dieser beachtliche Unterschied beruht möglicherweise darauf, dass die Applikation des Proteasominhibitors die Balance zwischen Proteinsynthese und Abbau massiv stört, was zu einer Zunahme der intrazellulären Proteinkonzentration führt und die Kapa-

zität, der für die Faltung der neu synthetisierten Proteine zuständigen Systeme, übersteigt[71]. Dadurch wird ein großer Anteil der neu synthetisierten Proteine fehlgefaltet und durch den fehlenden Abbau werden sie dem Pool der DRiPs zugeordnet. Weiterhin haben Vabulas und Hartl[92] gezeigt, dass die Bedingungen, unter denen die DRiP Rate bestimmt wurden, keine physiologischen sind und damit der Anteil der neu synthetisierten Proteine überschätzt wird. Bei den Markierungsexperimenten wurden die Zellen vor Verabreichung des Proteasominhibitors „gehungert“, d. h. in einem Medium ohne Aminosäuren präinkubiert und dann mit radioaktiv markierten Aminosäuren versorgt. Durch die Präinkubation ändern sich die interne Proteinausstattung und damit die Proteinsyntheserate nach Umsetzung in ein anderes Medium. Somit wird die DRiP Rate in diesen Experimenten überschätzt, da die gesamte Proteinsyntheserate steigt und dieser Anstieg fälschlich dem DRiP Pool zugeordnet wird.

Die hier vorgestellten quantitativen Analysen legen den Schluss nahe, dass eine DRiP Rate von etwa 10% plausibler ist, als eine Rate von mehr als 35%. Zum einen sind DRiPs „defektive“, also fehlerhaft, weichen somit in Struktur oder Sequenz von nativen Proteinen ab. Eine hohe DRiP Rate führt zu einer hohen Konzentration dieser Proteine, was das Risiko einer Zellschädigung mit sich bringt, da die für die Faltung zuständige Maschine[14, 15, 28] damit überfordert wäre und diese Proteine zu „Inclusion Bodies“ aggregieren würden[27]. Zum anderen korreliert unter stationären Bedingungen, also beim balanciertem Generieren und Entfernen von MHC Class I Komplexen an der Zelloberfläche, die Menge der aus DRiPs stammenden Antigene mit der DRiP Rate, was bedeutet, dass die Selbsterkennung durch die T-Zellen empfindlich von der reproduzierbaren Herstellung der DRiPs abhängt. Hier stellt sich die Frage, ob es möglich ist, mehr als 35% der Selbsterkennung durch das Immunsystem einem Prozess zu überlassen, der als stochastisch und als Fehler des Proteinsynthesystems angesehen wird. Dieses würde dazu führen, dass entweder eine eigentlich gesunde Zelle zerstört wird oder das Immunsystem weniger sensitiv sein muss als ohne fehlerhaften Antigene.

Yewdell et al.[31] argumentieren, dass Fehler in der Produktion von Makromolekülen unvermeidlich sind und damit die von ihnen vermutete DRiP Rate von mehr als 35% begründet. Da aber Beispielsweise bei der DNA Replikation eine , durch energieabhängige Fehlerkorrektur, eine Fehlerquote von unter 10^{-9} erreicht werden kann[93], erscheint dieses Argument nicht stichhaltig. Dieses begründet die Hypothese, dass der schnelle Abbau einer substantiellen Fraktion der neu synthetisierten Proteine daraus herrührt, dass diese Proteine im hochkonzentrierten Milieu des Cytosols sich nicht schnell genug zu einem nativem Protein falten und daher abgebaut werden. Über Steuerung der Aktivität des Chaperonsystems könnte die Zelle sogar innerhalb gewisser Grenzen die Höhe der DRiP

Rate einstellen. Eine Reihe von Arbeiten belegen die Verknüpfung der Antigenpräsentation mit der Aktivität von Chaperonen[94, 95, 96].

2.4.2 DRiPs sind nützlich für die Immunüberwachung und beschleunigen die Präsentation metabolisch stabiler Proteine.

Die Zeitverläufe der Präsentation von SIINFEKL an der Zelloberfläche können ohne die Annahme eines DRiP Pools nicht beschrieben werden. Zusätzlich zu den beobachteten nativen Proteinen wird noch ein Proteinpool mit einer relativ kurzen Lebenszeit benötigt, jedoch sind 10% dafür ausreichend. Der Beitrag von DRiPs zur Antigenpräsentation besteht vor allem darin, dass langlebige und nicht cytosolische Proteine ebenfalls an der Zelloberfläche repräsentiert werden können (Abbildung 12). Ohne den sofortigen Abbau einer kleinen Fraktion aller neu synthetisierten Proteine könnten Viren metabolische Stabilität der eigenen Proteine als Strategie nutzen, der Immunüberwachung zu entgehen. Durch die Verkürzung der Zeitspanne zwischen Produktion eines Proteins und der Präsentation daraus entstandener Antigene bei langlebigen Proteinen, wozu viele virale Proteine zählen, erhält das Immunsystem einen entscheidenden Vorteil in der Bekämpfung viraler Infektionen beziehungsweise der Erkennung von unphysiologischen Veränderungen im Proteinhaushalt der Zellen (Abbildung 11). Durch die Erzeugung von DRiPs wird es möglich, den Herstellungsprozess der Proteine zu überwachen und nicht die Proteinausstattung der Zellen. Für kurzlebige cytosolische Proteine bringen DRiPs keine entscheidende Veränderung für die Immunüberwachung.

2.4.3 Das Verhältnis von abgebauten Proteinen zu präsentierten Antigenen ist im Mittel 1'000 zu 1 und ist wahrscheinlich gleich für DRiPs und native Proteine.

Entsprechend der Simulationen und in Übereinstimmung mit der Abschätzung von Reits et al.[24] müssen durchschnittlich 1'000 Proteine abgebaut werden, um ein SIINFEKL Peptid an der Zelloberfläche zu präsentieren. Dieser Wert für die Präsentationseffizienz ($\eta \sim 10^{-3}$) liegt zwischen den berichteten Werten von 10^{-2} [91] und 10^{-4} [90]. Diese niedrigen Effizienzen beruhen auf einer geringen Ausbeute bei der Herstellung der Epitope oder deren N-terminal verlängerten Vorläufern durch das Proteasom, dem schnellen Abbau von Peptiden im Cytosol und ER sowie auf der Selektivität beziehungsweise Affinität der verschiedenen Schritte involviert im Weiterleiten der Peptide zur Zelloberfläche. Cascio et al.[97] haben gezeigt, dass 6-8% aller proteasomal abgebauten Ovalbumin Proteine ein SIINFEKL Epitop oder ein N-terminal verlängertes Peptid produzieren. Darüber hinaus werden Peptide sehr effizient im Cytosol durch Aminopeptidasen mit einer Halbwertszeit

von 2-3 Sekunden abgebaut[88]. Die cytosolische Endopeptidase Thimet hat anscheinend eine besondere Rolle beim Abbau von proteasomal hergestellten potentiellen Antigenen[62]. Auch wenn eine kleine Fraktion von Peptiden im Cytosol durch Chaperone geschützt wird[94], werden die meisten zerstört, bevor sie das ER erreichen[88]. Auch im ER sind Aminopeptidasen aktiv, die Peptide abbauen können, bevor sie an MHC Komplexe binden[64, 65]. Wenn man das hohe Risiko eines Peptids betrachtet, zerstört zu werden, bevor es an der Zelloberfläche präsentiert wird, ist es wahrscheinlich, dass die Herstellung N-terminal verlängerter Vorläufer ein Mechanismus ist, das im Peptid enthaltene Epitop vor Aminopeptidasen zu schützen[87, 98].

Eine wesentliche Annahme des hier entwickelten Modells ist die gleiche Effizienz der Antigenpräsentation für Epitope, die aus den Pools der DRiPs und der nativen Proteine stammen. Begründet wird dieses zum einen mit der Überlegung, dass, wenn das Protein einmal vom 19S Komplex des 26S Proteasoms entfaltet worden ist und sich im inneren des Proteasoms befindet, die Information, aus welchem Pool es stammt, verloren ist[99, 100]. Es ist allerdings möglich das unterschiedliche Effizienzen bestehen, da defektive Proteine akkumulieren können und dabei spezielle subzelluläre Kompartimente bilden[27, 71]. Mit dem hier beschriebenen Modell kann man den Effekt unterschiedlicher Effizienzen der beiden Pools simulieren. Hierzu wurde das Verhältnis der Präsentationseffizienzen von DRiP zu nativen Proteinen einmal auf 0.1 und einmal auf 10 festgelegt und die anderen freien Parameter des Modells an die experimentellen Daten angepasst. Auch hier konnten die Kurven aus Abbildung 9 reproduziert werden, jedoch ändert sich die DRiP Rate, also der Anteil der DRiPs an der gesamten Proteinsynthese. Bei einer hohen Effizienz des DRiP Pools sinkt die DRiP Rate, so haben die drei Konstrukte NP-SIINFEKL-GFP, KEKE-NP-SIINFEKL-GFP und R-NP-SIINFEKL-GFP Raten von 1.2, 1.2 und 1.0%. Wenn man andersherum eine niedrige Effizienz annimmt, so steigt die DRiP Rate auf 22, 35 und 47%. Der Beitrag zur Antigenpräsentation (Abbildung 10) bleibt allerdings in etwa gleich. Die Modelle sind hinsichtlich ihrer Güte nicht unterscheidbar, höchstens über Plausibilitätsbetrachtung der erhaltenen Parameter, wie sie im Ergebnisteil dieses Kapitels angestellt wurde. Festzuhalten bleibt aber, dass es keinen Grund für eine niedrige Effizienz der Präsentation von Antigenen aus dem DRiP Pool gibt, wie die Überlegungen im vorhergehenden Abschnitt zeigen.

2.4.4 Hinweise auf eine wechselseitige Koordination von Proteinsynthese und Proteinabbau.

Die festgestellte Reduktion des proteasomalen Proteinabbaus auf 35% des Ausgangswertes bei Hemmung der Proteinsynthese durch Cycloheximid weist auf die Existenz eines

zellulären Regulationsmechanismus zur Aufrechterhaltung des zellulären Protein Pools hin. Offensichtlich reduziert eine Senkung des Proteinabbaus die Verringerung der Proteinmengen bei Inhibition oder kompletten Blockierung der Proteinsynthese. Es gibt ältere Arbeiten[81, 101], die in Bakterien und in tierischen Zellen zeigen, dass die Proteolyserate sinkt, wenn die Synthese inhibiert wird, vor allem bei geringer Nährstoffversorgung. Der Effekt von Cycloheximid auf Synthese und Abbau von Proteinen wurde schon 1969 beschrieben[79]. Die Studien[77, 83] zeigen einen Einfluss von Cycloheximid auf den lysosomalen Proteinabbau. Neuere Studien zeigen eine substratspezifische Wirkung von Cycloheximid[78]. So führt Inhibition der Proteinsynthese durch Cycloheximid zu einer Akkumulation von polyubiquitinierten murine μ Ketten in *S. cerevisiae* Hefe aber der Abbau der Carboxypeptidase Y wurde nicht verändert. Der Einfluss von Cycloheximid auf den Proteinabbau weist auf einen Sensor hin, der den Zustand der Proteinsynthese an das Proteasomsystem vermittelt. Die Art dieses Sensors ist bisher unbekannt, so wurde beispielsweise vermutet, dass der Sensor ein Protein ist, welches permanent produziert und vom Proteasom abgebaut wird[80] aber nachgewiesen wurde dieses nicht. Vabulas und Hartl[92] zeigten, dass bei Hemmung des proteasomalen Proteinabbaus weniger Proteinsynthese durchgeführt wird, da weniger Aminosäuren als Bausteine zur Verfügung stehen.

Zusammengenommen gibt es einige experimenteller Hinweise, die auf eine Verknüpfung und gegenseitige Beeinflussung der Systeme für Synthese und Abbau von Proteinen hinweisen. Dieses hat weit reichende Konsequenzen, wenn man Ergebnisse von Studien analysiert, die die Proteinsynthese hemmen, um Proteinabbau und Antigenproduktion zu analysieren. Beispielsweise führt die Inhibition der Proteinsynthese innerhalb kurzer Zeit zu einer signifikanten Zunahme der Beweglichkeit des TAP Transporters[24]. Eine höhere Beweglichkeit des Peptidtransporters weist auf eine verringerte Beladung hin, was als Fehlen von Peptiden aus DRiPs als Substrat für TAP interpretiert wurde. Die Schlussfolgerung war, dass neu synthetisierte Proteine einen Großteil der TAP Substrate und damit der präsentierten Antigene stellen. Wenn man jedoch eine Verringerung der proteasomalen Aktivität annimmt, wird die Gesamtmenge der für TAP zur Verfügung stehenden Peptide gesenkt und somit ebenfalls eine signifikante Erhöhung der Beweglichkeit von TAP bewirkt. Auch wenn die hier dargestellte Arbeit einen Beitrag zur Diskussion dieses Punktes beigetragen hat, so ist die Kontroverse über den Beitrag der DRiPs zur Antigenpräsentation noch immer nicht abgeschlossen[29, 76].

2.4.5 Ist das Modell zu stark vereinfacht, um verlässliche Aussagen über die Rolle der DRiPs in der Antigenpräsentation abzuleiten?

Die hier vorgestellten Schlussfolgerungen basieren auf Parameterwerten, die aus einem relativ einfachen kinetischen Modell stammen. Man könnte befürchten, dass die Einfachheit und Unvollständigkeit des Modells die Verlässlichkeit der Ergebnisse beschränkt. Dieser Überlegung kann man jedoch entgegen, dass beschränktes beziehungsweise unvollständiges Wissen über die zugrunde liegenden molekularen Prozesse immer ein Problem bei der Interpretation von systemischen Daten darstellt unabhängig davon, ob die Interpretation auf einem mathematischen Modell beruht. Zum anderen heißt „einfach“ für ein Modell nicht notwendigerweise ungeeignet oder falsch[102]. Das Zusammenfassen von individuellen Reaktionen zu einer überschaubaren Zahl von übergreifenden Reaktionen, mit interpretierbaren und gegebenenfalls messbaren Parametern und das Beschreiben ihres kinetischen Verhaltens durch phänomenologische Ratengleichungen bringt häufig eine bessere quantitative Beschreibung und Verständnis des Systemverhaltens als sehr detaillierte kinetische Modelle mit vielen Freiheitsgraden und Ungewissheiten bezüglich der vielen individuellen Reaktionen und Parameter[103, 104].

Die hier durchgeführten Untersuchungen wurden, wie auch die experimentellen Studien, an einzelnen Beispielen durchgeführt. Somit sind die erhaltenen Ergebnisse und Parameterwerte möglicherweise abhängig von den jeweiligen betrachteten antigenenthaltenden Proteinen sowie den im Experiment verwendeten Zelltypen.

Festzuhalten bleibt noch, dass sich der Abbau der Proteine hier mit einem Ratengesetz erster Ordnung ausreichend gut beschreiben lässt, solange der Abbau der spezifischen Proteine nicht zentral gesteuert wird.

Die hier dargestellten Ergebnisse konnten im Journal of Immunology publiziert werden[105].

3 Vorhersage von potentiellen Epitopen aus der Proteinsequenz

In diesem Kapitel wird die Frage erörtert, welche proteolytischen Fragmente eines Proteins als Antigen an der Zelloberfläche präsentiert werden. Es werden Methoden vorgestellt, die es erlauben, aus der Sequenz eines Proteins die wahrscheinlich präsentierten Epitope abzuleiten.

Ziel dieser Arbeit ist zum einen möglichst gute Vorhersagen zu machen, aber zum anderen ist es ebenfalls Ziel, die einzelnen zellulären Systeme, die im Prozessieren der Antigene involviert sind, besser zu verstehen beziehungsweise die Prinzipien abzuleiten, nach denen sie funktionieren. So werden hier keine „über alles“ Modelle entwickelt, wie beispielsweise die SYFPEITHI Datenbank[106], sondern jeweils individuelle Modelle für die einzelnen Schritte des MHC Class I Pathways, die dann gegebenenfalls zusätzlich für die bessere Vorhersage von Antigenen kombiniert werden.

Begriffsdefinitionen

Vorhersagemethode: Die gesuchte Eigenschaft Y eines Objekts soll aus den bekannten Eigenschaften X abgeleitet werden. Dazu wird mittels eines Trainingsdatensatzes von n Objekten, bei denen Y ebenfalls bekannt ist, eine Funktion f gebildet, so dass der Ausdruck:

(8)

$$\sum_{j=1}^n \text{norm}(Y^j - f(X^j, P))$$

minimal ist. P sind hierbei die Parameter der Vorhersagemethode f und norm beschreibt den Abstand zwischen der Vorhersage f(X,P) und der bekannten Eigenschaft Y. Es muss sowohl die Form von f als auch die dazugehörigen Parameter bestimmt werden. Die Funktion f ist die Vorhersagemethode.

Overfitting: Overfitting ist die Überanpassung der Parameter eines Modells an den Trainingsdatensatz. Die Vorhersagequalität für einen unabhängigen Datensatz ist vermindert. Ursache sind zu viele Freiheitsgrade beziehungsweise Parameter im Modell im Vergleich zu den zur Verfügung stehenden Daten. Ein gutes prädiktives Modell ist der Kompromiss zwischen einem übermäßig einfachen Modell, das sowohl den Trainingsdatensatz schlecht reproduzieren kann, als auch Vorhersagen geringer Qualität für unabhängige Datensätze

macht und einem Modell, das aufgrund der vielen Freiheitsgrade lediglich den Trainingsdatensatz „auswendig lernt“ und somit ebenfalls schlechte Vorhersagen macht.

3.1 Peptid Bindung an MHC Class I Moleküle

Die Bindung der Peptide an das MHC I Molekül und die Bildung von MHC Komplexen im ER wird als der am stärksten selektive Schritt im Pathway bis zur Präsentation an der Zelloberfläche angesehen[107, 108]. Es existieren mehrere Modelle zur Vorhersage der Bindung der Peptide an die MHC-Moleküle, wobei zu beachten ist, dass für jedes MHC I Allel, von denen allein für den Menschen mehr als 2000 bekannt sind, ein eigenes Modell gebildet werden muss. Eine aktuelle Gegenüberstellung mehrerer Methoden findet man in der Arbeit von Peters et al. [109].

Die im weiteren Verlauf verwendete „stabilisierte Matrix Methode“ (SMM) [110] und andere Methoden zum erstellen von Vorhersagemodellen, werden im Anhang dieser Arbeit erläutert.

3.1.1 Verwendete Modelle für MHC Class I Allele

In dieser Arbeit wird kein neues Modell für die Bindung von Peptiden an MHC I Moleküle erstellt, sondern bestehende Modelle genutzt.

Für das HLA A0201 Allel ist eine matrixbasierte Vorhersage für Affinitäten von 9mer Peptiden zum MHC Molekül von Peters et al. entwickelt worden[111]. Für das Training der Matrix nach der SMM Methode sind 533 Peptide verwendet worden, die keine Epitope darstellen. Diese Matrix ist in der Lage, Epitope mit einer hohen Qualität von Zufallspeptiden zu unterscheiden. Der AUC⁵[112] Wert auf dem HLA A0201 Datensatz ist 0.919.

Für die Maus MHC Allele H2 K^b, H2 L^d und H2 D^b werden Matrizen verwendet, die von Udaka et al. veröffentlicht wurden[113]. Die Anzahl der bekannten und in SYFPEITHI enthaltenen Epitope liegt für die einzelnen Allele zwischen 9 und 21. Für das H2 Kb Allel ist sowohl eine Matrix für 9mer Peptide vorhanden, als auch eine Matrix für 8mer Peptide. Um die letztere Matrix verwenden zu können, ist ein weiterer Datensatz (H2 Kb8) erstellt worden. Er beinhaltet alle in der SYFPEITHI Datenbank enthaltenen 8mer Epitope des Allels und die zugehörigen Zufallspeptide.

Die auf diesen Matrizen basierende Vorhersage von MHC Molekül Affinitäten für Peptide wird in dieser Arbeit als MHC Modell bezeichnet.

⁵ Area Under Curve ist die Fläche unter der Kurve bei der Auftragung von Sensitivität gegen 1-Spezifität einer Vorhersagemethode. Der Wert 1 wäre eine perfekte Methode, 0.5 eine zufällige Vorhersage.

Die IEDB[114, 115] ist eine weitere geeignete Quelle von Vorhersagemodellen für die MHC I Molekülbindung für eine Vielzahl an MHC Allelen in verschiedenen Spezies.

3.2 TAP Transport

Der Transport von Peptiden aus dem Cytosol in das ER mittels TAP ist proportional zur Bindung der zu transportierenden Peptide[116, 117, 118]. Daher wurde im Rahmen der Modellentwicklung und vor allem in Hinsicht auf das Ziel der Vorhersage von potentiellen Epitopen die Affinität der Peptide zu TAP betrachtet und vorhergesagt.

3.2.1 Verwendete experimentelle Daten

Hier wird eine Übersicht über die in der Arbeit verwendeten Daten und ihre Herkunft gegeben.

In vitro Experimente zur Bestimmung der Affinität von Peptiden zu TAP

Die Affinitätsdaten bestehen aus IC_{50} Werten und den entsprechenden Peptidsequenzen beziehungsweise aus Matrizen, die Beiträge zu IC_{50} Werte beinhalten. Ein ausführlich beschriebener, experimenteller Assay zum Bestimmen des IC_{50} Wertes ist beispielsweise in der Arbeit von Uebel et al.[117] gegeben.

9mer Peptid Affinitäten

Der erste Datensatz beinhaltet Affinitätswerte für 478 9mer Peptide[119]. Von diesen sind nur 430 verwendet worden, da für 26 Peptide der Affinitätswert nicht exakt bestimmt ist. Bei diesen weggelassenen Peptiden ist die Affinität sehr niedrig, so dass man mit sehr hohen und nicht mehr sinnvollen Konzentrationen hätte arbeiten müssen. Die anderen 22 Peptide sind bekannte T- Zell Epitope, deren Verwendung möglicherweise die Untersuchungen die Vorhersagekapazität des Modells zur Identifikation potentieller Antigene verfälschen würde.

Unter der Annahme des unabhängigen Beitrags der Aminosäuren eines Peptids zur Bindung an TAP wurde mittels SMM eine Affinitätsmatrix generiert. Sie ist in Tabelle 3 dargestellt, wobei die Werte auf das Polyalaninpeptid normiert wurden.

Tabelle 3: SMM Matrix für TAP vermittelten Peptidtransport

	(N1) Pos 1	(N2) Pos 2	(N3) Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8	(C) Pos 9	
A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	A
C	1.96	0.32	-0.05	-0.38	0.19	-0.30	-0.39	-0.26	-0.33	C
D	2.17	1.36	0.69	-0.92	0.34	-0.40	0.12	0.81	0.28	D
E	2.66	0.40	0.79	-0.54	0.32	-0.91	0.22	0.27	1.43	E
F	2.62	-0.07	-0.90	-0.74	0.20	-0.34	-0.98	-0.11	-1.92	F
G	2.29	1.18	0.08	-0.95	0.50	0.05	0.57	0.23	0.37	G
H	1.89	0.35	-0.48	-0.05	-0.41	-0.16	-0.40	0.20	-0.11	H
I	1.99	0.62	0.04	-0.89	-0.30	-0.73	-1.30	0.73	-0.61	I
K	1.16	-0.11	0.36	-0.84	-0.54	-0.58	-0.16	-0.20	-0.32	K
L	0.98	1.09	0.42	-0.55	-0.32	-0.55	-1.15	0.43	-1.26	L
M	1.79	0.35	-0.24	-0.80	-0.29	-0.42	-0.77	0.39	-0.47	M
N	0.89	0.62	1.03	-1.04	-0.04	-0.47	-0.04	0.63	0.23	N
P	3.02	1.85	0.70	-0.83	-0.62	-0.55	-0.28	0.43	-0.38	P
Q	2.37	-0.24	0.20	-0.68	0.16	-0.14	0.01	0.19	0.24	Q
R	0.95	-1.42	-0.19	-1.01	-0.44	-0.89	-0.37	-0.86	-1.08	R
S	1.35	0.55	0.30	-0.52	0.02	-0.34	-0.12	0.07	0.58	S
T	1.82	0.35	0.34	-1.13	-0.13	-0.20	-0.10	0.23	0.10	T
V	1.72	-0.19	-0.11	-0.72	0.37	-0.87	-0.45	0.18	0.01	V
W	2.15	-0.03	-1.53	-0.82	-0.76	-0.74	-1.42	-0.60	-0.44	W
Y	2.45	-0.19	-1.43	-0.99	-0.07	-0.60	-1.00	-0.14	-3.09	Y

Die Tabelle gibt die logarithmierten Beiträge (natürlicher Logarithmus) der einzelnen Aminosäuren an den angegebenen Positionen zum IC₅₀ Wert des Peptids an.

Affinitäten von längeren Peptiden

Der nächste Datensatz besteht aus 64 längeren Peptiden[120] und stammt ebenfalls von Professor Peter van Endert. Die Längenverteilung der Peptide in diesem Datensatz ist in Abbildung 13 dargestellt. Dieser Datensatz ist zu einem sehr viel späteren Zeitpunkt übermittelt worden als der zuvor beschriebene Datensatz. Er ist nicht für die Modellbildung verwendet worden und stellt somit einen unabhängigen Testdatensatz dar.

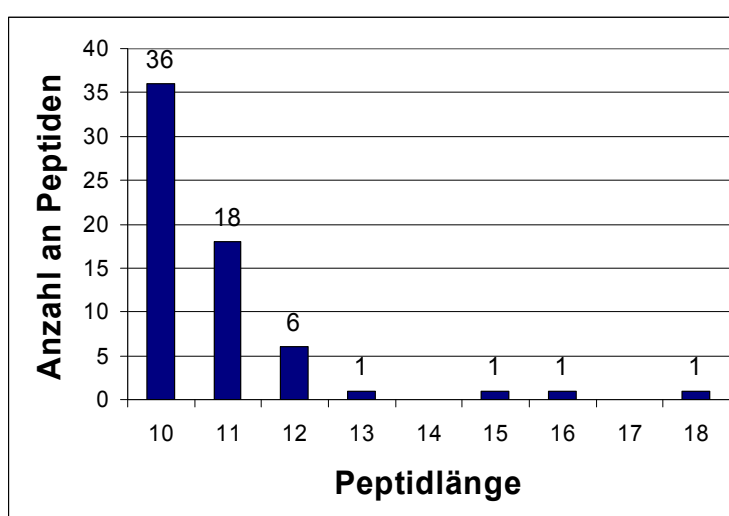


Abbildung 13: Längenverteilung, des aus 64 längeren Peptiden bestehenden Datensatzes

Auf der x-Achse ist die Peptidlänge angegeben und auf der y-Achse die Anzahl der Peptide für die jeweilige Länge. Die Anzahl an Peptiden ist sehr ungleich auf diesem Datensatz verteilt.

Experimentell bestimmte Austauschmatrizen zur TAP Affinität

Es existierten zwei experimentelle Austauschmatrizen. Die in ihnen enthaltenen Daten sollen ebenfalls in die hier erstellte Vorhersagemethode einfließen.

Die erste Matrix[36, 121] basiert auf zufällig zusammengestellten Peptidbibliotheken und wird hier als Mix Matrix bezeichnet (Tabelle 4). Die Beiträge der einzelnen Positionen sind auf die Affinität einer komplett randomisierten Peptidbibliothek normiert.

Tabelle 4: Mix Matrix für TAP vermittelten Peptidtransport

	(N1) Pos 1	(N2) Pos 2	(N3) Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8	(C) Pos 9	
A	-0.44	0.12	-0.64	-0.47	-0.37	0.17	0.69	0.07	0.81	A
C	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	C
D	1.54	0.72	1.71	-0.60	0.08	0.68	1.05	0.24	2.02	D
E	1.70	0.01	1.81	-0.74	0.08	0.19	1.02	0.45	1.77	E
F	1.49	-0.59	-0.20	-0.78	-0.55	-0.26	0.07	0.22	-1.92	F
G	0.39	0.66	1.85	0.29	-0.03	0.42	0.61	-0.10	0.99	G
H	-0.01	-0.65	0.11	-0.55	-0.41	-0.35	-0.49	0.28	0.60	H
I	0.00	-1.03	-0.36	-0.53	-0.39	-0.42	-0.94	0.24	-0.48	I
K	-1.30	-0.40	-0.27	-0.13	-0.33	-0.10	-0.03	-0.21	-0.17	K
L	-0.78	0.94	-0.57	-0.32	0.14	-0.08	-0.70	-0.16	-1.25	L
M	-0.48	-0.38	-0.22	-0.91	0.01	0.15	-0.55	-0.52	-0.39	M
N	-1.68	0.50	0.70	1.20	0.25	-0.15	-0.07	0.37	1.52	N
P	1.24	2.96	-0.17	0.31	-0.68	-0.45	-0.80	-0.03	-0.31	P
Q	-0.04	-1.03	-0.01	0.53	0.21	0.02	-0.11	-0.04	0.04	Q
R	-1.32	-1.46	-0.71	0.22	-0.33	-0.91	-0.06	-0.60	-1.82	R
S	-0.66	-0.41	0.27	0.66	0.64	0.10	0.46	0.08	1.91	S
T	-0.19	0.25	0.40	0.71	1.03	0.53	0.33	-0.32	0.11	T
V	-0.70	0.21	-0.53	0.69	0.27	-0.18	-0.17	0.16	-0.78	V
W	0.99	0.24	-1.36	0.28	-0.04	0.61	-0.03	-0.28	-0.55	W
Y	0.25	-0.66	-1.80	0.15	0.42	0.02	-0.27	0.14	-2.11	Y

Die Tabelle gibt die logarithmierten Beiträge (natürlicher Logarithmus) der einzelnen Aminosäuren an den angegebenen Positionen zum IC₅₀ Wert des Peptids an.

Die zweite Matrix[122, 123] basiert auf dem Austausch von einzelnen Aminosäuren des alaninreichen Peptids AAASAAAAY und wird als Ala Matrix bezeichnet.

Jede Stelle des Grundlagenpeptides wird gegen jede Aminosäure ausgetauscht, wobei immer nur eine einzige Aminosäure ausgetauscht wird und man erhält wieder eine Matrix mit 171 Einträgen, da Cystein auch hier nicht verwendet werden kann. Die Einträge sind auf das Peptid AAAAAAAAY normiert.

Tabelle 5: Ala Matrix für TAP vermittelten Peptidtransport

	(N1) Pos 1	(N2) Pos 2	(N3) Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8	(C) Pos 9	
A	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.36	A
C	2.44	0.54	-0.35	-0.46	0.03	0.13	0.29	-0.05	3.86	C
D	4.64	3.06	2.73	-0.37	0.51	0.34	1.91	0.00	6.72	D
E	4.85	0.53	1.58	0.32	1.16	0.00	0.56	0.69	5.06	E
F	3.22	0.18	-2.41	-1.17	-0.51	0.51	-0.69	0.49	-0.20	F
G	2.42	2.48	2.81	0.82	1.44	-0.45	2.94	-1.20	6.40	G
H	3.00	2.17	-0.65	-1.20	0.44	0.00	0.21	0.76	4.67	H
I	1.92	-0.16	-1.90	-0.04	-0.63	-1.43	-0.69	0.47	3.06	I
K	1.31	0.17	-0.16	-0.92	0.59	-0.45	1.39	0.85	2.65	K
L	2.55	-0.89	-0.54	0.02	-0.92	0.36	-0.46	-0.15	3.19	L
M	1.79	-0.46	-1.61	-0.53	-0.54	0.82	-0.69	-0.13	3.51	M
N	0.73	1.82	0.94	-0.21	1.20	-0.51	0.49	0.06	5.75	N
P	4.28	5.08	-0.22	-0.80	-0.92	0.29	-1.56	-0.25	3.96	P
Q	3.33	-0.76	0.65	-0.58	0.03	-0.42	1.01	0.69	3.61	Q
R	0.59	-0.65	-0.69	-1.20	-0.25	-0.78	0.03	0.29	2.01	R
S	1.88	-0.27	-0.58	-0.51	0.63	0.85	2.16	-1.61	7.82	S
T	2.27	0.16	0.21	-0.06	0.34	0.80	0.83	-1.20	5.47	T
V	1.76	-0.65	-1.83	-0.36	0.43	0.66	-0.37	0.03	3.40	V
W	2.73	-1.24	-2.41	-1.17	-1.61	0.72	-0.15	-0.92	1.90	W
Y	3.03	-0.27	-2.53	-0.87	-0.80	1.10	-1.47	0.14	0.00	Y

Die Tabelle gibt die logarithmierten Beiträge (natürlicher Logarithmus) der einzelnen Aminosäuren an den angegebenen Positionen zum IC₅₀ Wert des Peptids an.

In sowohl in der Mix Matrix als auch in der Ala Matrix wird für Cystein der Mittelwert aller Aminosäuren an der entsprechenden Position verwendet, da diese Werte, aufgrund der Schwefelbrückenbindung bei hohen Cystein Konzentrationen, mittels Peptidbibliotheken nicht bestimmbar sind.

In vivo beobachtete T Zell Epitope

In der SYFPEITHI Datenbank[106] (<http://www.syfpeithi.de>) ist eine Vielzahl von T- Zell Epitopen aufgelistet, zusammen mit der Proteinsequenz, aus der sie stammen und sortiert nach dem MHC Allel, das sie präsentiert.

Es wurden sämtliche 9mer Epitope von Mensch und von Maus aus der Datenbank verwendet, für die die Sequenz, aus der sie stammen, bekannt war und die genau einmal in dieser Sequenz vorkommen. Es wurden keine Epitope aus Signalpeptiden verwendet und auch nicht die in der Datenbank ebenfalls abgelegten Beispiele für MHC Liganden, da diese nicht von TAP transportiert worden sind. Die aus SYFPEITHI extrahierten Daten wurden in drei Gruppen geteilt. So erhält man für die Maus den H2-X Datensatz, der 67 9mer Epitope aus 56 Sequenzen beinhaltet. Für den Menschen wurden zwei Datensätze erstellt. Diese sind zum einen der HLA A0201 Datensatz, der aus 87 9mer Epitopen aus 51 Sequenzen besteht, die vom HLA A0201 Allel präsentiert werden, sowie der HLA-X Datensatz mit 203 9meren aus 87 Sequenzen, die alle von anderen menschlichen MHC

Class I Allelen präsentiert werden. Alle 9mer Peptide, die man aus den zu den Epitopen gehörenden Sequenzen gewinnen kann und die nicht das Epitop sind, wurden als Zufallsneunmere bezeichnet und gehören zu dem entsprechenden Datensatz. Für die von menschlichen MHC Molekülen präsentierten Epitope sind zwei Datensätze erstellt worden, weil zum einen das HLA A0201 Allel sehr ausgiebig untersucht worden ist und daher einen Gesamtdatensatz zu stark beeinflussen würde. Zum andern ist eine MHC Affinitätsvorhersage für dieses Allel von Peters et al. entwickelt worden [111] und somit kann auf diesem Datensatz die Kombination von TAP und MHC getestet werden. Auch kann man durch die Trennung der Datensätze das Modell auf einem Datensatz entwickeln und dann auf den anderen Datensätzen testen.

Die SYFPEITHI Datenbank stellt eine Sammlung von gefundenen T Zell Epitopen dar und wird permanent erweitert. In einem späteren Abschnitt dieser Arbeit wurde eine neuere Version der Datenbank verwendet und somit zusätzliche Epitope und Sequenzen. Dieses ändert aber nicht die Verwendbarkeit des hier vorgestellten Datensatzes.

3.2.2 Beschreibung der Peptidaffinität zu TAP

Aus dem Datensatz der 430 9mer Peptide und deren Affinität zu TAP wurde mit der SM-Methode eine weitere Matrix erstellt (Tabelle 3). Somit stehen drei unterschiedlich erstellte Matrizen zur Verfügung, um TAP Affinitäten von Peptiden vorherzusagen. In Abbildung 14 wurden die Ergebnisse der Vorhersagemethoden für den 9mer Peptid Datensatz dargestellt. Hierbei ist zu beachten, dass die Vorhersage durch die SM-Methode mittels 5-facher Kreuzvalidierung durchgeführt wurde, da ja der gleiche Datensatz zum Training der Matrix verwendet worden ist.

Die Mix Matrix macht bessere Vorhersagen als die Ala Matrix, was in dieser Form zu erwarten ist, da durch den Einsatz von Peptidbibliotheken mehr Effekte und gegenseitige Beeinflussung der Aminosäuren in einem Peptid erfasst werden, als beim Einsatz des konstanten Peptids bei der Ala Matrix. Die SMM Matrix macht noch einmal deutlich bessere Vorhersagen für die Affinität von Peptiden zu TAP. Es ist hier jedoch darauf hinzuweisen, dass trotz der Kreuzvalidierung die Vorhersage nicht gleichwertig zu einem völlig unabhängigen Datensatz betrachtet werden kann, da einzelne Gruppen von im Datensatz enthaltenen Peptiden zueinander sehr ähnlich sind. In Abbildung 15 ist die Vorhersage der SMM Matrix für den Trainingsdatensatz ohne Kreuzvalidierung dargestellt. Die Güte ist für den Trainingsdatensatz deutlich besser.

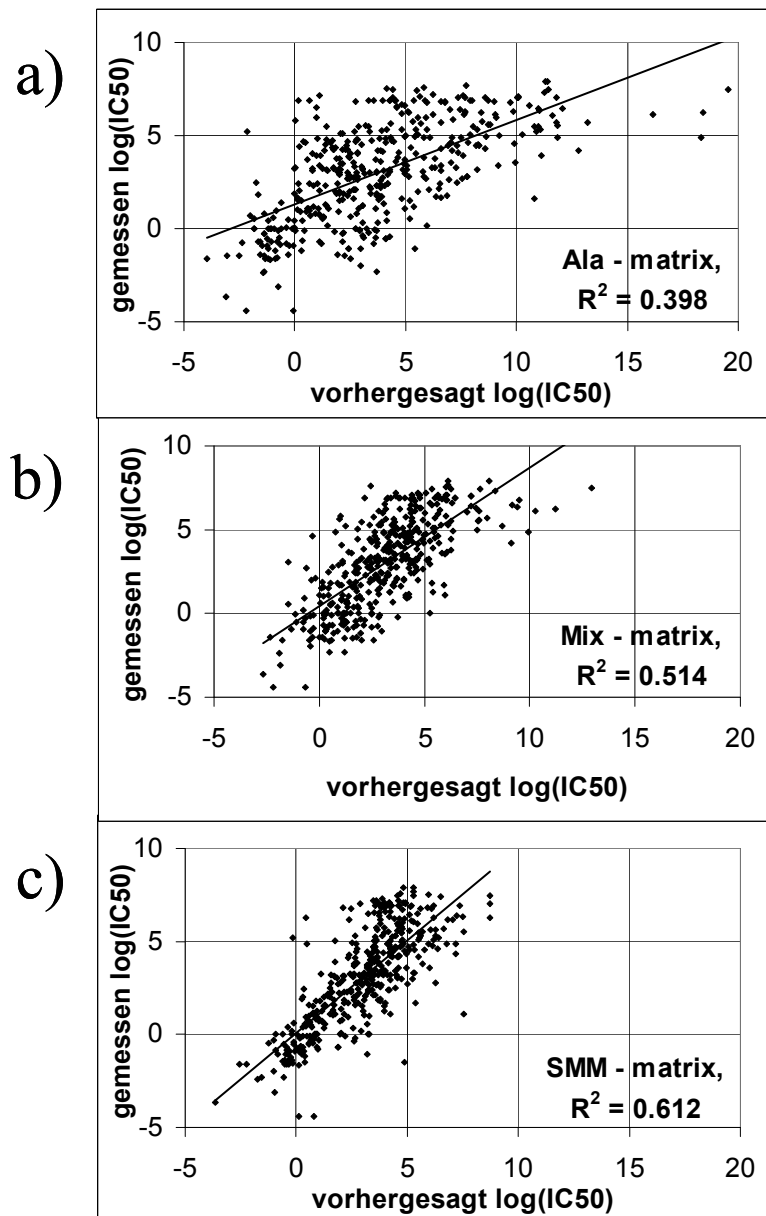


Abbildung 14: Scatterplots der TAP Affinität für 9mer Peptide experimentell und vorhergesagt

In den drei Diagrammen ist für jedes Peptid des Datensatzes der experimentell gemessene IC50 Wert gegen den vorhergesagten aufgetragen. Die IC50 Werte sind logarithmiert. Die Vorhersage wird mit unterschiedlichen Matrizen durchgeführt: a) Ala- Matrix b) Mix- Matrix c) SMM- Matrix. Die Vorhersage der SMM- Matrix ist 5-fach kreuzvalidiert. Die Korrelation zwischen gemessenen und vorhergesagten Werten ist angegeben (R^2).

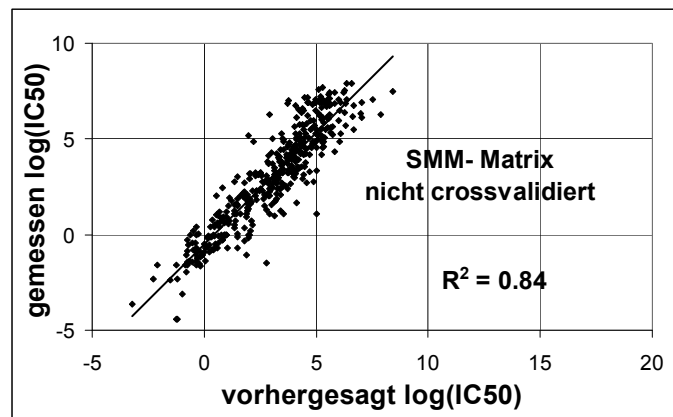


Abbildung 15: Scatterplot der TAP Affinität für 9mer Peptide experimentell und SMM Matrix nicht kreuzvalidiert

Die Auftragung ist identisch mit Abbildung 14. Die SMM- Matrix Affinitätsvorhersage wird nicht kreuzvalidiert.

Die drei verwendeten Matrizen sind aus unterschiedlichen Datensätzen entstanden und sind auf unterschiedliche Art „verrauscht“. Es wurde eine vierte Matrix aus den vorhergehenden drei Matrizen gebildet, indem die drei Matrizen gemittelt worden sind. Jeder Eintrag der neuen Matrix ist das arithmetische Mittel der drei zugehörigen Einträge der anderen Matrizen. Die neue Matrix heißt Konsensmatrix.

Auch für diese Matrix wurden die Affinitäten der 9mer Peptide vorhergesagt. Die Konsensmatrix beschreibt die Affinitäten geringfügig besser als die SMM Matrix (Abbildung 16). Auch die Vorhersage mittels der Konsensmatrix auf dem 9mer Peptid Affinitäten Datensatz wurde 5-fach kreuzvalidiert durchgeführt, da in der Konsensmatrix die SMM- Matrix enthalten ist. Trotz der geringen Verbesserung der Vorhersagequalität gegenüber der SMM Matrix ist die Konsensmatrix als deutlich besser anzusehen, da Peptide, die durch den Trainingsdatensatz schlecht abgedeckt wurden durch die zusätzliche Information der Austauschmatrizen besser beschrieben werden können. Dieses wird bei der Anwendung der Methoden auf einen unabhängigen Datensatz in einem späteren Abschnitt dieser Arbeit deutlich (Abbildung 19).

Tabelle 6: Konsensmatrix für TAP vermittelten Peptidtransport

	(N1) Pos 1	(N2) Pos 2	(N3) Pos 3	Pos 4	Pos 5	Pos 6	Pos 7	Pos 8	(C) Pos 9	
A	-1.56	-0.25	-0.10	0.24	-0.10	0.17	0.27	0.00	0.55	A
C	0.05	-0.01	-0.02	0.11	0.09	0.05	0.00	-0.13	0.00	C
D	1.37	1.42	1.83	-0.23	0.33	0.32	1.07	0.32	1.83	D
E	1.65	0.02	1.51	0.08	0.54	-0.13	0.64	0.44	1.58	E
F	1.03	-0.45	-1.05	-0.50	-0.26	0.08	-0.50	0.17	-2.52	F
G	0.28	1.14	1.70	0.45	0.66	0.12	1.41	-0.38	1.41	G
H	0.21	0.33	-0.23	-0.21	-0.11	-0.06	-0.19	0.39	0.55	H
I	-0.11	-0.49	-0.62	-0.09	-0.42	-0.75	-0.94	0.45	-0.52	I
K	-1.03	-0.41	0.09	-0.23	-0.08	-0.26	0.44	0.12	-0.45	K
L	-0.50	0.09	-0.11	0.11	-0.34	0.02	-0.73	0.01	-0.94	L
M	-0.38	-0.46	-0.58	-0.35	-0.26	0.30	-0.64	-0.11	-0.29	M
N	-1.43	0.69	1.01	0.38	0.49	-0.27	0.16	0.33	1.33	N
P	1.43	3.00	0.22	-0.04	-0.72	-0.13	-0.84	0.03	-0.09	P
Q	0.47	-0.97	0.39	0.15	0.15	-0.07	0.34	0.26	0.12	Q
R	-1.34	-1.47	-0.42	-0.27	-0.32	-0.75	-0.09	-0.42	-1.47	R
S	-0.56	-0.34	0.11	0.27	0.45	0.31	0.87	-0.51	2.26	S
T	-0.12	-0.04	0.43	0.23	0.43	0.49	0.39	-0.46	0.72	T
V	-0.49	-0.50	-0.71	0.27	0.37	-0.02	-0.29	0.10	-0.30	V
W	0.54	-0.64	-1.65	-0.18	-0.78	0.31	-0.50	-0.63	-0.87	W
Y	0.50	-0.67	-1.80	-0.18	-0.13	0.28	-0.87	0.02	-2.91	Y

In jedem Feld ist der Beitrag einer Aminosäure zur Affinität als relativer logarithmierter IC50 Wert an der jeweiligen Position angegeben. Farblich hervorgehoben sind Beiträge, die die Affinität stark erhöhen (rot), erhöhen (orange), verringern (grün) oder stark verringern (blau).

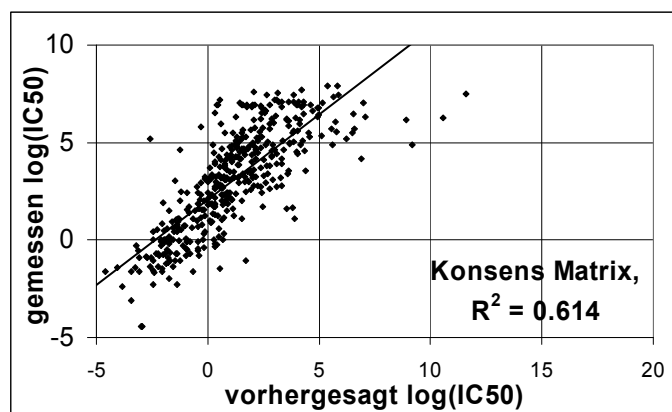


Abbildung 16: Scatterplot der TAP Affinität für 9mer Peptide experimentell und Konsensmatrix

Die Auftragung ist identisch mit Abbildung 13. Die Vorhersagemethode ist die Konsensmatrix. Es ist wie bei der SMM- Matrix 5-fach kreuzvalidiert.

In der Konsensmatrix (Tabelle 6) sind die Aminosäuren mit den stärksten Einflüssen auf die Affinität des Peptids hervorgehoben. Die Positionen des Peptids, bei denen die Beiträge der Matrix am stärksten vom Mittelwert abweichen, haben den stärksten Einfluss auf die Bindung des Peptids. Dieses ist noch einmal in Abbildung 17 dargestellt. Man sieht, dass die N-terminalen und die C-terminalen Position stärkeren Einfluss auf die Affinität des Peptids ausüben als die Positionen in der Mitte des Peptids. Es ist klar zu erkennen, dass an der C-terminalen Position und an den drei N-terminalen Positionen gehäuft Aminosäuren auftreten, die einen starken Einfluss auf die Affinität des Peptids ausüben.

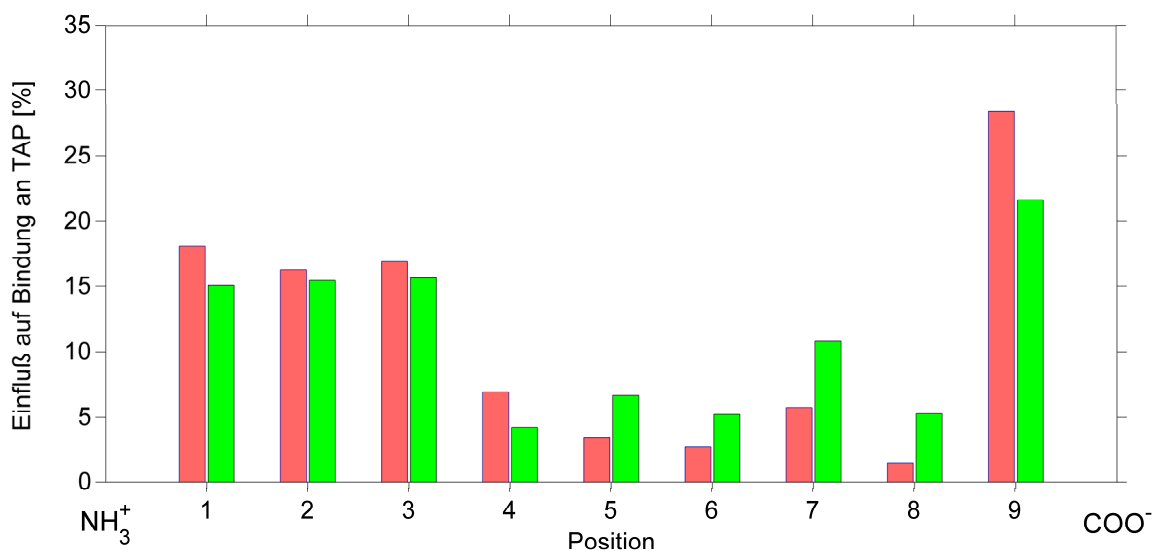


Abbildung 17: Einfluss der Position auf die Affinität des Peptids

Beitrag der einzelnen Positionen zur Bindung an TAP. Grün: Konsensmatrix, die Ergebnisse stimmen qualitativ mit den Ergebnissen in der Arbeit von Uebel et al. [124] überein.

Tabelle 7: Einteilung der Aminosäuren

Position	1	2	3	4	5	6	7	8	9
Konsensmatrix									
erhöhen Affinität	A,R,K	Q,R	F,W,Y				I		F,L,R,W,Y
vermindern Affinität	D,E,F,P	D,G,P	D,E,G,N				D,G		D,E,G,N,S
Uebel et al.									
erhöhen Affinität	K,N,R	R,I,Q	W,Y			R	I		F,L,R,Y,V
vermindern Affinität	D,E,F	P,L	D,E,G	N	T				D,E,N,S,G

Beim Vergleich zwischen den Ergebnissen der Konsensmatrix und den von Uebel et al. [124] berichteten in Tabelle 7, treten Unterschiede auf. Zum Teil lassen sich dieses auf die Willkür der Einteilung erklären, da es keine eindeutige Vorschrift gibt, ab welchem

Beitrag eine Aminosäure günstig, ungünstig oder neutral ist. Zum anderen wird in dieser Arbeit darauf Wert gelegt, das experimentelle Rauschen zu vermindern, wodurch leicht unterschiedliche Resultate zu erwarten sind. Insgesamt kann man jedoch sagen, dass die Ergebnisse weitgehend übereinstimmen. So gibt es keine Aminosäure, die in einer Einteilung einen günstigen und in der anderen Einteilung einen ungünstigen Einfluss ausübt.

3.2.3 Erweiterung des TAP Modells mit Paarkoeffizienten

Das oben entwickelte Matrixmodell hat als Grundannahme, dass die Bindung einer Aminosäure an TAP keinen Einfluss auf die Bindung einer anderen Aminosäure hat. Diese Annahme kann eingeschränkt werden, wenn man berücksichtigt, dass eine Aminosäure an einer Position die Affinität einer anderen Aminosäure verändern kann. Wenn Aminosäure 1 an Position 1 Aminosäure 2 an Position 2 beeinflusst, muss Aminosäure 2 Aminosäure 1 genauso beeinflussen, da man den Effekt nicht einzeln beobachten kann. Man erhält so für 9mer Peptide $9 \cdot 20 \cdot 8 \cdot 20 / 2 = 14400$ Paarkoeffizienten und es ist nicht möglich, Aussagen über alle Paarkoeffizienten aus der geringen Anzahl an Beobachtungen zu gewinnen. Wie in der Arbeit von Peters et al.[111] beschrieben, kann man nur Aussagen über einzelne Paarkoeffizienten machen, für die man über genügend Beobachtungen verfügt, da sonst das Problem des „Overfitting“ noch stärker zur Geltung kommt und die Vorhersagequalität für unabhängige Daten viel schlechter wird. Man kann die Paarkoeffizienten auf verschiedene Arten interpretieren. Zum Beispiel können durch die Bindung einer Aminosäure der Sequenz Spannungen im Peptid erzeugt werden, sich das Target durch „induced fit“ etwas ändern, die Peptide Sekundärstrukturmerkmale ausprägen, sich Ladungen der Seitenketten gegenseitig beeinflussen oder ein sequentieller Mechanismus vorliegen.

Hier wurden Paarkoeffizienten auf der Basis der Mix-Matrix, der SMM Matrix und der Konsens Matrix erstellt und mit dem 9mer Peptid Affinitäten Datensatz trainiert.

In Abbildung 18 ist die 5-fach kreuzvalidierte Vorhersage der Konsensmatrix mit Paarkoeffizienten auf den 9mer Peptid Affinitäten Datensatz dargestellt. Die Tabelle 8 zeigt die durch Einführung der Paarkoeffizienten erreichte Verbesserung der einzelnen Methoden.

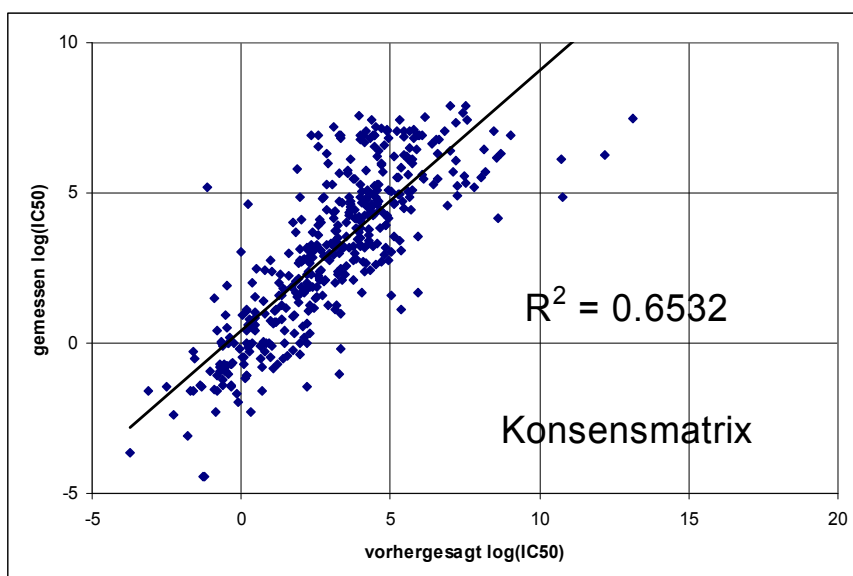


Abbildung 18: Scatterplot Vorhersage der Affinitäten der 430 9mere mit Konsensmatrix und Paarkoeffizienten

Auftragung wie in Abbildung 19. Es wird eine fünffach kreuzvalidierte Vorhersage mit der Konsensmatrix und den Paarkoeffizienten durchgeführt und nur Paarkoeffizienten verwendet, für die mindestens 10 Beobachtungen im Datensatz vorhanden sind.

Tabelle 8: Vorhersagequalität mit und ohne Paarkoeffizienten

Matrix	R ² ohne Paarkoeffizienten	R ² mit Paarkoeffizienten
Mix	0.514	0.576
SMM	0.612	0.616
Konsens	0.614	0.653

Es wurde noch einmal eine deutliche Verbesserung durch Einführen der Paarkoeffizienten erreicht. Beim Training der SMM Matrix ist die gleiche Datenmenge verwendet worden wie beim Training der Paarkoeffizienten. Da nur Paarkoeffizienten bestimmt werden können, die häufig in dieser Datenmenge vorkommen, werden diese Paarkoeffizienten teilweise schon beim Training der SMM Matrix vorweggenommen. Die Häufung verschiedener Kombinationen von Aminosäuren in den Peptiden des Trainingsdatensatzes führt zu einem Bias in der Vorhersagematrix. Methodisch korrekter ist es, die Matrix aus zufälligen Peptidbibliotheken, wie die Mix- Matrix, zusammenzustellen und dann die Paarkoeffizienten auf einem unabhängigen Datensatz zu trainieren. Leider beschreibt die Mix- Matrix die Peptide allein noch nicht gut genug, so reicht auch die große Verbesserung durch die Paarkoeffizienten nicht aus, um bessere Vorhersagen zu erhalten, als die SMM Matrix liefert. Die Konsensmatrix berücksichtigt die meisten Daten und sollte durch die Mittlung am wenigsten stark durch experimentelles „Rauschen“ belastet sein. Auch ist hier der

Einfluss der SMM Matrix nicht so groß, dass keine Paarkoeffizienten mehr unabhängig (oder nahezu unabhängig) ermittelt werden können. Das Erweitern der Methode mit Paarkoeffizienten führt zu Verbesserungen. Die beste Methode, um die Affinitäten von 9mer Peptiden vorherzusagen, ist die Konsensmatrix kombiniert mit Paarkoeffizienten.

Die geringe Verbesserung der SMM Matrix mit Paarkoeffizienten zeigt, dass fast die gesamte Information, die im Trainingsdatensatz enthalten ist, unter Annahme von unabhängigen Beiträgen zur Affinität, wiedergeben kann. Die Verbesserung der Konsensmatrix mit Paarkoeffizienten ist möglicherweise eine Anpassung der Matrix an den Trainingsdatensatz. Das Problem des „Overfittings“ ist bei Paarkoeffizienten deutlich größer und die Anzahl an Beobachtungen ist nicht so hoch, dass das auszuschließen ist.

Es ist festzustellen, dass die Annahme der unabhängigen Bindung für TAP akzeptabel ist, da die Vorhersagequalität der Matrix Methoden gut ist und die Erweiterung mit Paarkoeffizienten nicht deutlich besser ist und möglicherweise zum Teil eine Anpassung der unabhängigen Matrizen an den Trainingsdatensatz darstellt.

3.2.4 Erweitern des TAP Modells für längere Peptide

Die matrixbasierten Vorhersagen haben den Nachteil, dass sie nur auf Peptide mit einer festen Länge (hier 9 Aminosäuren) angewendet werden können. Da TAP aber auch Peptide anderer Länge transportiert, müsste man für jede Länge von Peptiden einen ausreichend großen Datensatz aufnehmen und damit separat eine Matrixmethode aufbauen. Diese Affinitätsdaten für längere Peptide existieren leider nur in geringer Menge (Abbildung 13).

Hier wird die Tatsache genutzt, dass die drei N-terminalen Aminosäuren und der C-Terminus des Peptids (N1, N2, N3, C) die Affinität des Peptids am stärksten bestimmen (Abbildung 17). Für längere und kürzere Peptide wird ein Transportscore t gebildet, der nur diese vier Matrixeinträge berücksichtigt.

(9)

$$t = mat_{1,N1} + mat_{2,N2} + mat_{3,N3} + mat_{9,C}$$

Der erste Index von $mat_{1,N1}$ bezeichnet die Position in der Matrix und der zweite Index bezeichnet die Aminosäure des Peptids, die zu bewerten ist. Peptide mit einer Länge von 8 bis 16 Aminosäuren werden, im Gegensatz zu längeren und kürzeren Peptiden, mit annähernd gleicher Affinität gebunden[125], so dass kein Term in das Modell eingearbeitet werden muss, der die Länge des Peptids berücksichtigt, wenn man nur Peptide in diesem Längenbereich betrachtet.

Man kann die Methode auf längere Peptide anwenden. In Abbildung 19 sind Scatterplots der Vorhersage von Affinitäten für den Datensatz der 64 längeren Peptide dargestellt. Auf diesem unabhängigen Datensatz wird deutlich, dass die Konsensmatrix tatsächlich eine Verbesserung zur SMM Matrix darstellt. Die Korrelation zwischen vorhergesagten und experimentellen Werten dieses komplett unabhängigen Datensatzes ist nicht ganz so hoch wie bei der Beschreibung der 9mer Peptide, jedoch werden auch die Affinitäten dieser längeren Peptide gut wiedergegeben.

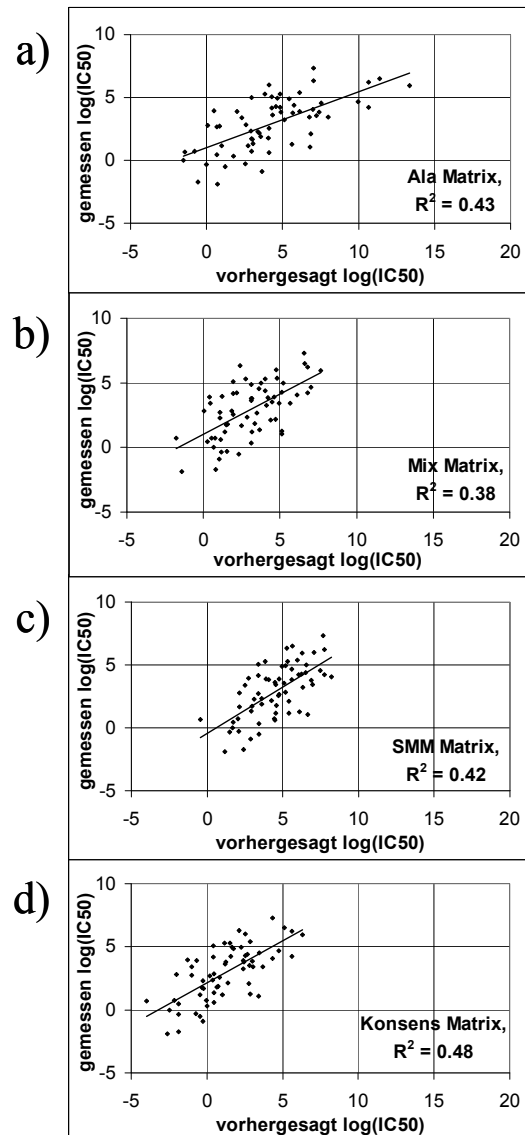


Abbildung 19: Scatterplots der TAP Affinität von 64 längeren Peptiden experimentell und vorhergesagt

Die Auftragung ist wie in Abbildung 13. Es werden vier verschiedene Matrizen zur Vorhersage verwendet a) Ala- Matrix, b) Mix- Matrix, c) SMM- Matrix und d) Konsensmatrix.

Der große Vorteil dieser Methode ist, dass sie auf Peptide unterschiedlicher Länge angewendet werden kann und somit auf alle potentiellen Substrate von TAP. Erst so wird es möglich den Einfluss von TAP auf die MHC I Antigen Präsentation zu untersuchen, was im nächsten Abschnitt dieser Arbeit beschrieben wird.

3.2.5 Einfluss von TAP auf die Antigenpräsentation

Welchen Einfluss hat TAP im MHC Class I Pathway auf die Selektion der Epitope? Kann man anhand der Funktion des Transporters zwischen Epitopen und Nichtepitopen unterscheiden? Wie hoch ist die Selektivität von TAP? Diese Fragen werden in diesem Abschnitt untersucht.

Da die Affinität sehr eng mit der Transportrate zusammenhängt[116], kann hier ein Affinitätsmodell verwendet werden, um Transportprozesse zu beschreiben.

Vorhersage von Epitopen mit der Konsensmatrix

Als erster Test werden für die Epitope und Zufallsneunmere des HLA A0201 Datensatzes mit dem 9mer Matrixmodell von TAP Affinitäten vorhergesagt. Die Ergebnisse werden in Klassen eingeteilt. Man kann in Abbildung 20 sehen und auch mit einem χ^2 -Test (17 Freiheitsgrade, Irrtumswahrscheinlichkeit 5%) prüfen, dass die Verteilungen der TAP Affinitäten von Epitopen und Zufallsneunmeren sich signifikant unterscheiden. Die Epitope haben leicht höhere Affinitäten (=niedrigere IC_{50} Werte). Bemerkenswert ist außerdem, dass sich die TAP Affinitäten der Epitope über mehrere Größenordnungen erstrecken.

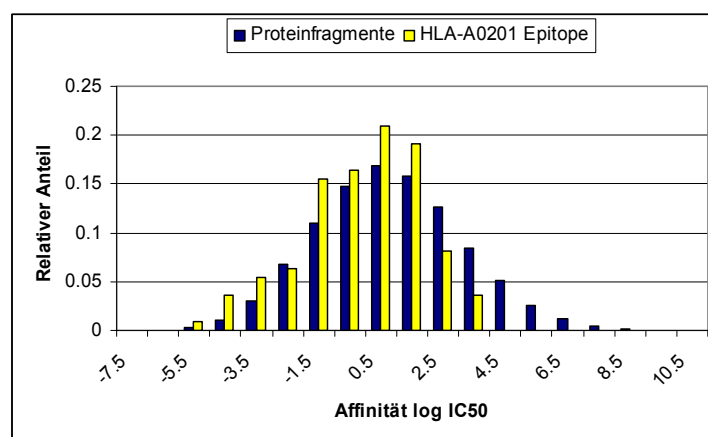


Abbildung 20: Verteilung der TAP Affinitäten für HLA-A0201 Datensatz

Auf der x-Achse ist die Affinität zu TAP in logarithmierten IC_{50} Werten aufgetragen. Die Epitope und die Proteinfragmente werden nach ihren vorhergesagten Affinitäten in Klassen eingeteilt. Auf der y-Achse ist der relative Anteil einer Klasse an der Gesamtmenge an Peptidsequenzen aufgetragen

Um zu beurteilen, wie gut man mit der TAP Affinitätsvorhersage zwischen Epitopen und Zufallsneunmeren unterscheiden kann, wird eine ROC⁶[112] Kurve dargestellt (Abbildung 21) und der AUC- Wert berechnet (AUC= 0.611).

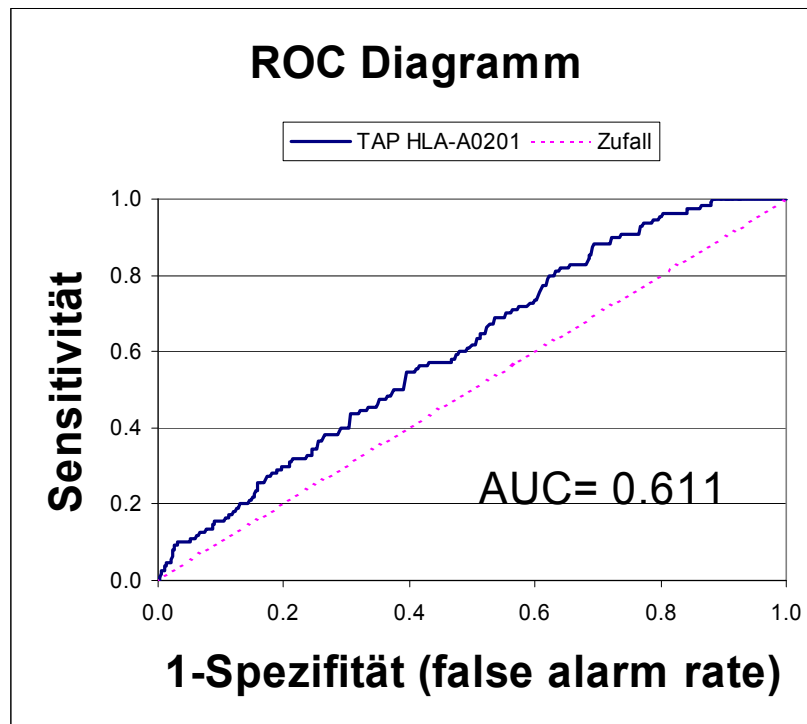


Abbildung 21: ROC Diagramm der Epitopvorhersage für HLA-A0201 mit der TAP Konsensmatrix

Auf der x-Achse ist 1-Spezifität aufgetragen und auf der y-Achse die Sensitivität. Der Cutoff-Wert, bei dem entschieden wird, ob ein Peptid Epitop ist oder nicht, ist der Laufparameter entlang der Kurve. Eine völlig zufällige Trennung wäre die Diagonale (rot). Die Trennung für den HLA-A0201 Datensatz ist in blau dargestellt.

Wenn man die TAP Affinitätsvorhersage mit einer MHC Bindungsvorhersage [111] kombiniert, wird die sehr gute Vorhersage von Epitopen, die die MHC Bindungsvorhersage erreicht, durch die TAP Vorhersage verschlechtert (Daten nicht gezeigt).

Dieser erste Test hat gezeigt, dass TAP offenbar einen Einfluss auf die Selektion von Epitopen hat, aber die Vorhersage in dieser Art nicht ausreicht, um die Trennung zwischen Epitopen und Zufallsneunmeren anhand der sehr guten MHC Affinitätsvorhersage zu verbessern.

⁶ Receiver Operating Characteristic ist die Auftragung von Sensitivität (=Trefferquote; [positiv erkannt]/[tatsächlich positiv]) gegen die Fehlalarmrate (1-Spezifität (= [negativ erkannt]/[tatsächlich negativ])) einer Vorhersagemethode. Die Art der Darstellung ist auf dem Gebiet des maschinellen Lernens verbreitet.

Einbeziehen von Epitopvorläufern

An dieser Stelle wird das entwickelte rein statistische Modell durch mechanistische Annahmen über das Prozessieren von Epitopen erweitert. Ausgegangen wird von dem allgemeinen Modell für die Affinität von Peptiden zu TAP, mit dem sich nicht nur 9mere, sondern auch längere Peptide bewerten lassen.

Als erstes wird wieder untersucht, wie gut man die Epitope des HLA-X Datensatzes von den zugehörigen Zufallsneunmeren durch TAP Affinitäten trennen kann. Man erhält einen AUC Wert von 0.700. Dieser AUC Wert ist nur geringfügig kleiner als der AUC Wert von 0.702, den man erhielt, wenn man die komplette 9mer Matrix, anstatt der verkürzten Matrix, die nur die 3 N-terminalen und die C-terminale Aminosäure bewertet, verwenden würde. Der MHC HLA-0201 Datensatz soll hier für die Modellentwicklung nicht verwendet werden, da in einem späteren Abschnitt auf diesem Datensatz überprüft wird, wie sich MHC Bindungsvorhersage und TAP Affinitätsvorhersage kombinieren lassen.

Es wird angenommen, dass N-terminal verlängerte Vorläuferpeptide von TAP in das ER transportiert werden. Die Bewertung der 9mer Peptidsequenzen muss dafür verändert werden. Man kennt die Sequenz, aus der das 9mer Peptid stammt und kennt daher auch die Sequenz eines N-terminal verlängerten Vorläuferpeptids. Für dieses Vorläuferpeptid erhält man eine Reihe von möglichen Affinitätswerten, da nicht bekannt ist, welcher Vorläufer transportiert wird. Beispielsweise erhält man für ein 9mer Peptid insgesamt 4 Affinitätswerte (9mer; 10mer; 11mer und 12mer), wenn man Vorläufer der maximalen Länge von 12 Aminosäuren betrachtet. Wie kann man aus diesen Werten einen vernünftigen Wert für das Peptid bilden?

(i) Da keine anderen Informationen vorhanden sind, wird angenommen, dass für das betrachtete Peptid alle Vorläuferpeptide bis zur betrachteten Länge vom Proteasom in gleicher Menge erstellt werden. Dann ergibt sich für das Peptid ein Affinitätswert nach folgender Rechnung:

Ausgehend von den IC_{50} Werten wird zuerst die transportierte Menge eines Peptids berechnet. Zu beachten ist, das zwischen IC_{50} Werten (=Verhältnis zur Referenzkonzentration) und den logarithmierten Werten (=Beitrag zur Bindungsenergie) unterschieden wird.

(10)

$$Menge = k \frac{1}{IC_{50}}$$

Die Menge des transportierten Peptids ist proportional zur Affinität des Peptids. Der Pro-

portionalitätsfaktor k entspricht hierbei der transportierten Menge eines hypothetischen Peptids mit einem IC_{50} Wert von 1 (nicht logarithmiert). Die Gesamtmenge an transportierten Peptiden, die ein Epitop bilden können, ergibt sich aus der Summe der Einzelmengen.

(11)

$$Gesamtmenge = k * \sum_{i=l_E}^L \frac{1}{IC_{50,i}}$$

L ist die maximale Länge der Vorläuferpeptide und die Länge des Epitops l_E ist hier gleich 9. Aus dieser Gesamtmenge kann ein effektiver IC_{50} Wert durch Einsetzen in (10) abgeleitet werden.

(12)

$$\log(\text{effektiver } IC_{50}) = \log \left(\frac{1}{\frac{1}{\sum_{i=l_E}^L IC_{50,i}}} \right)$$

(ii) Die gleiche Annahme wie unter (i), aber zusätzlich wird angenommen, dass TAP nur den Vorläufer mit der höchsten Affinität transportiert beziehungsweise die Unterschiede in den Transportraten so groß sind, dass alle außer dem besttransportierten Peptid vernachlässigt werden können. Aus Gleichung 11 erkennt man, dass das Vorläuferpeptid mit dem niedrigsten IC_{50} Wert die Gesamttransportmenge am stärksten beeinflusst. Man erhält so:

(13)

$$\text{effektiver } IC_{50} = \min_{i=l_E}^L (IC_{50,i})$$

(iii) Die Annahme ist, dass das Proteasom nur ein Vorläuferpeptid herstellt, aber nicht bekannt ist, welches tatsächlich existiert, und dass alle hypothetischen Vorläuferpeptide mit der gleichen Wahrscheinlichkeit hergestellt werden können. Die beste Schätzung ist dann der Mittelwert der Affinitäten.

(14)

$$\begin{aligned}\bar{t}_L &= \frac{1}{L-l_E+1} \sum_{i=l_E}^L \log(IC_{50,i}) \\ &= mat_{l_E,C} + \frac{1}{L-l_E+1} + \sum_{i=l_E}^L mat_{1,N1} + mat_{2,N2} + mat_{3,N3}\end{aligned}$$

Hier wird die dritte Variante verwendet, da sie am besten auf den Datensätzen funktioniert und auch sehr einfach zu verwenden ist. Die Unterschiede zwischen den verschiedenen Möglichkeiten der Bewertung der Peptide sind sehr gering.

AUC Werte werden für verschiedene Längen von Vorläuferpeptiden auf dem HLA-X Datensatz bestimmt.

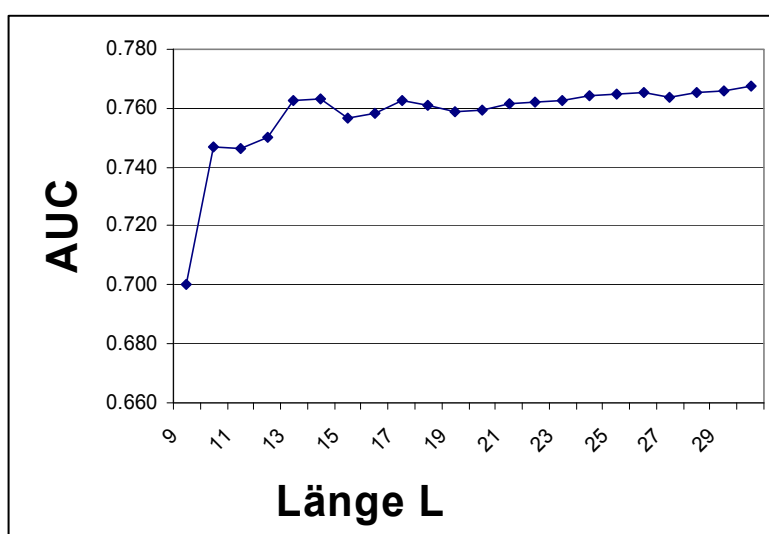


Abbildung 22: AUC der TAP Vorhersage über betrachtete Länge des Peptids für HLA-X

Auf der x-Achse ist die Länge des Epitop Vorläuferpeptids aufgetragen. So wird bei Länge 9 nur das 9mer betrachtet, bei Länge 10 das 9mer und das 10mer usw. Auf der y-Achse ist der AUC Wert aufgetragen. Der AUC Wert steigt mit zunehmender Länge des betrachteten Peptids sogar über die zu erwartende physiologische Länge von Peptiden hinaus an.

Wie man in Abbildung 22 sehen kann, verbessert sich die Trennung von Epitopen und Zufallsneunmeren, wenn man längere Peptide als Vorläufer akzeptiert. Es ist nicht zu erwarten, dass man einen weiteren Anstieg der AUC Werte bei Vorläuferpeptiden sehen kann, die länger als 18 Aminosäuren sind, da TAP diese Peptide mit einer geringeren Effizienz transportiert.

Dieses interessante Ergebnis wirft die Frage auf, bei welcher Länge des Vorläuferpeptids man keine Verbesserung des AUC Wertes mehr erfährt. Ein hypothetischer Endwert ist

die Qualität, die durch alleiniges Verwenden des C-Terminus erreicht werden würde, da bei der Berücksichtigung von hypothetisch unendlich langen Vorläufern über alle verschiedenen N-Termini gemittelt wird und so die N-Termini für alle Peptide gleich sind. Die Trennung durch den C-Terminus führt zu einem AUC Wert von 0.782, der deutlich höher ist als die in Abbildung 22 erreichten AUC Werte.

Weiterentwicklung des Modells und Einführen des Parameters α zum Gewichten des N-Terminus

Die erste untersuchte Hypothese ist, dass der Transport keinen Gleichgewichtsprozess darstellt, sondern Peptide sofort transportiert werden, sobald sie an TAP gebunden haben und nicht miteinander um die Bindungsstelle an TAP konkurrieren. Wenn man dazu annimmt, dass das Peptid zuerst mit dem C-Terminus an TAP bindet, kann man die vermuten, dass der C-Terminus *in vivo* einen sehr viel höheren Einfluss hat als *in vitro*.

Als einfaches Modell ist dafür die Annahme gemacht worden, dass der C-Terminus des betrachteten Peptids eine Mindestaffinität zu TAP haben muss, damit der N-Terminus überhaupt gebunden werden kann. Diese Mindestaffinität wurde im Modell systematisch zwischen den höchsten und niedrigsten möglichen Beiträgen des C-Terminus zur Affinität variiert. Auch der Ansatz, dass das Peptid transportiert wird, falls der C-Terminus allein eine ausreichend hohe Affinität zu TAP hat oder falls die Gesamtaffinität des Peptids ausreichend groß ist, wurde untersucht. Auch hier wurden die Parameter variiert. Der erste Ansatz stellt eine logische UND Verknüpfung zwischen der Affinität des C-Terminus und der Affinität des N-Terminus dar, während der zweite Ansatz einer logischen ODER Verknüpfung entspricht. Es sind einige Modellrechnungen durchgeführt worden, die aber alle zu Verschlechterungen der Trennung zwischen Epitopen und Zufallspeptiden geführt haben (Daten nicht gezeigt).

Diese Ergebnisse führten zu der Frage, ob der N-Terminus *in vivo* überhaupt einen Einfluss auf den Transport des Peptids in das ER hat oder nicht.

Es wurde deshalb ein Parameter α , zum Wichten des N-Terminus des Peptids beim Bewerten, eingeführt.

(15)

$$\begin{aligned}\bar{t}_L &= \frac{1}{L - l_E + 1} \sum_{i=l_E}^L \log(IC_{50,i}) \\ &= mat_{l_E,C} + \frac{\alpha}{L - l_E + 1} + \sum_{i=l_E}^L mat_{1,N1} + mat_{2,N2} + mat_{3,N3}\end{aligned}$$

Der Parameter α wird variiert und falls beim Übergang von α gleich 0 (Abbildung 23 blaue Kurve) nach α gleich 1 (Abbildung 23 gelbe Kurve) kein Maximum zwischendurch erreicht wird, kann man keinen selektiven Einfluss des N-Terminus erkennen und die Steigung des AUC für längere Vorläuferpeptide in Abbildung 22 ist nur ein Artefakt. Andernfalls hat der N-Terminus einen gewissen Einfluss. Das Optimum ist α gleich 0.2 (Abbildung 23 rote Kurve) und der zugehörige AUC Wert (0.791) ist signifikant höher als der für α gleich 0. Hier sieht man wie in Abbildung 22 eine Erhöhung des AUC Wertes mit steigender Länge des betrachteten Vorläufers und zusätzlich den zu erwartenden Abfall für große Längen des Peptidvorläufers gegen den AUC Wert der C-terminalen Trennung zwischen Epitopen und Zufallsneunmeren.

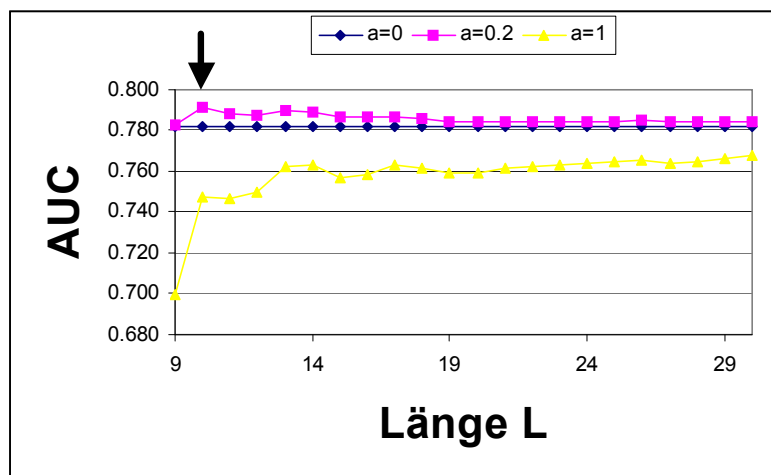


Abbildung 23: AUC der TAP Vorhersage über betrachtete Länge des Peptids für verschiedene α für HLA-X

Auftragung wie in Abbildung 22. Es wird der Verlauf für drei verschiedene α Werte dargestellt. Die beste Vorhersagequalität für alle α und alle Längen ist hervorgehoben. Der maximale AUC Wert ist 0.791.

Man kommt zu dem Ergebnis, dass *in vivo* N-terminal verlängerte Vorläufer von Epitopen transportiert werden, und dass der C-Terminus einen erheblich größeren Einfluss bei der Selektion von Epitopen durch TAP ausübt als auf die *in vitro* Affinität. In einem späteren Abschnitt wird sich zeigen, dass der große Einfluss des C-Terminus auf die Präferenz des Proteasoms beim Erstellen der Fragmente zurückzuführen ist.

Testen des Modells auf einem unabhängigen Datensatz

Da das Modell auf dem HLA-X Datensatz trainiert worden ist, muss man auf den anderen Datensätzen testen, ob qualitativ das gleiche Verhalten gezeigt wird. TAP von Maus unterscheidet sich in einigen Charakteristika von menschlichem TAP[126, 127]. So trans-

portiert menschliches TAP Peptide mit hydrophoben und basischen C-Termini, aber Maus TAP nur Peptide mit hydrophoben C-Termini. Es existierten zum Zeitpunkt dieser Untersuchungen nicht genügend Daten, um auch für Maus TAP eine Vorhersagematrix zu entwickeln, weshalb auch für den H2-X Datensatz die menschliche TAP Matrix verwendet wird. Auf diesem unabhängigen Datensatz soll nur getestet werden, ob man qualitativ die gleichen Ergebnisse erhält, also die besten Vorhersagen für α kleiner als 1 und beim Einbeziehen von Vorläuferpeptiden erzielt.

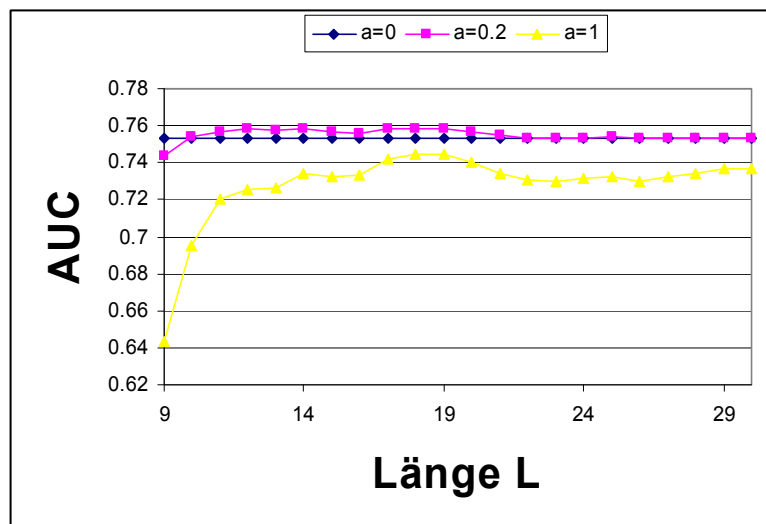


Abbildung 24: AUC der humanen TAP über Länge für murines MHC H2-X

Auftragung wie in Abbildung 22.

Wie man in Abbildung 24 erkennen kann, erhält man auf dem Maus Datensatz H2-X mit der für menschliches TAP entwickelten Methode insgesamt schlechtere Vorhersagen. Jedoch zeigen sich qualitativ die gleichen Ergebnisse wie für den HLA-X Datensatz. So kann geschlussfolgert werden, dass α kleiner als 1 eine Eigenschaft des TAP Transports widerspiegelt und keine Eigenschaft des HLA-X Datensatzes ist. Auf diesem Datensatz verbessern sich die Vorhersagen, wenn längere Vorläuferpeptide betrachtet werden jedoch nur geringfügig und das optimale α ist nochmals deutlich geringer als für den HLA-X Datensatz.

Mittlerweile existiert ein murines TAP Modell[126] und es bestätigt sich, dass der N-Terminus der Peptide für die Affinität des murinen TAP nahezu bedeutungslos ist.

Untersuchung der Rolle von TAP bei verschiedenen MHC Allelen

Wie stark variieren diese Ergebnisse zwischen den einzelnen MHC Allelen? Um diese Frage zu untersuchen, wird der HLA-X Datensatz in die individuellen MHC Allele aufgespalten.

Es werden nur Allele betrachtet, für die mindestens 10 Epitope bekannt und im Datensatz enthalten sind. Für jedes Allel einzeln wird die oben beschriebene Rechnung durchgeführt. Zuerst wird geprüft, wie gut man mit TAP Epitope von Zufallspeptiden trennen kann, ohne den N-Terminus mit α zu wichten oder längere Epitopvorläufer in betracht zu ziehen. Die Ergebnisse sind in Tabelle 9 Spalte 3 gezeigt. Es ist klar zu erkennen, dass es sehr große Unterschiede zwischen den verschiedenen Allelen gibt. Die Unterschiede in der Vorhersagequalität korrelieren mit der in [122, 128] getroffenen Einteilung von MHC Allelen in Allele, die gut mit TAP beladen werden (HLA-B27, HLA-A3 und HLA-A24) und in Allele, die ineffizient für die Beladung durch TAP sind (HLA-B07, HLA-B08 und HLA-A0201). Wenn man die Rechnung wiederholt und die für den gesamten Datensatz optimalen Parameter α gleich 0.2 und maximale Länge des Peptidvorläufers L gleich 10 verwendet (Tabelle 9 Spalte 4), sind die AUC Werte in einem viel engeren Bereich, nämlich zwischen 0.70 und 0.88 anstatt zwischen 0.39 und 0.89. So kann die Einteilung der MHC Allele in die zwei Klassen der Literatur (effizient für Beladung durch TAP, ineffizient für Beladung durch TAP) nicht aufrechterhalten werden. Man kommt zu dem Ergebnis, dass alle MHC Allele annähernd gleich von TAP beliefert werden, aber einige Allele die transportierten 9mere bevorzugen und andere MHC Allele diese nicht akzeptieren, sondern darauf angewiesen sind, dass TAP längere Vorläufer in das ER transportiert und diese im ER von der Aminopeptidase ERAAP getrimmt werden.

Tabelle 9: AUC für TAP Vorhersagen für verschiedene MHC Allele einzeln

	# Epitopes	AUC L=9, $\alpha=1$	AUC L=10, $\alpha=0.2$	Optimal α for L=10
HLA-B35	10	0.39	0.80	0.0
HLA-B07	11	0.43	0.71	0.0
HLA-B08	10	0.69	0.80	0.0
HLA-B44	11	0.78	0.88	0.0
HLA-A24	37	0.81	0.87	1.0
HLA-A3	11	0.82	0.75	1.2
HLA-B27	20	0.89	0.77	4.0
HLA-A0201	87	0.65	0.70	0.4

Es wird ermittelt, welcher Wert für α optimal ist, um die bestmögliche Trennung zu erreichen, wenn man L gleich 10 festhält (Tabelle 9 Spalte 5). Es ist zu erkennen, dass das optimale α von dem betrachteten Allel abhängt und keine ausschließliche Eigenschaft des Transports durch TAP ist.

Wenn das Motiv, das das MHC Molekül am C-Terminus des Peptids akzeptiert, mit dem Bindungsmotiv von TAP für den C-Terminus übereinstimmt, aber der N-Terminus nicht, sollte α sehr niedrig sein und das MHC Allel eher den Transport von Vorläufern von Epitopen benötigen. Wenn sowohl der N- als auch der C-Terminus für TAP und MHC gut passen, sollte α einen mittleren Wert annehmen. Wenn nur das Motiv des N-Terminus gut für

MHC und TAP übereinstimmt, sollte α größer werden, da der N-Terminus im Vergleich zum C-Terminus stärker gewichtet wird. Hier sollten verlängerte Vorläuferpeptide eher seltener zum Epitop werden, da das Epitop ja schon durch TAP transportiert werden kann und ein verlängerter Vorläufer noch einmal ein günstiges Motiv am N-Terminus, diesmal für TAP, besitzen müsste.

Im Mittel ist für den HLA-X Datensatz ein α von 0.2 optimal. Das deutet darauf hin, dass der C-Terminus, der für effektiven TAP Transport benötigt wird, mit dem mittleren MHC Bindungsmotiv für den C-Terminus des Peptids stärker übereinstimmt als die Motive des N-Terminus. Ein Grund dafür kann in der Koevolution des C-terminalen Bindungsmotivs von TAP und MHC liegen. Der C-Terminus des Peptids ändert sich nicht zwischen TAP und dem MHC Molekül, während der N-Terminus noch durch Aminopeptidasen wie ERAAP[63, 64, 65] verändert werden kann. Da der C-Terminus der meisten Peptide direkt durch das Proteasom hergestellt wird, erscheint es klar, dass die an der Präsentation dieser Moleküle beteiligten Systeme eine ähnliche Präferenz für den C-Terminus aufweisen, wie das Proteasom.

3.2.6 Diskussion des humanen TAP Modells und des Einflusses von TAP auf die Antigenpräsentation

In diesem Kapitel ist eine Methode zur Vorhersage der humanen TAP Transporteffizienz für Peptide beliebiger Länge entwickelt worden. Um dieses zu erreichen, wird genutzt, dass die drei N-terminalen und die C-terminalen Aminosäure eines Peptids den größten Einfluss auf die Affinität des Peptids zu TAP haben und dass die Affinität des Peptides zum Transporter proportional zur Transportrate dieses Peptids ist. So werden die Affinitätsmatrizen, die anhand von experimentell bestimmten 9mer Peptid Bindungsstärken zu TAP entwickelt worden sind, verwendet, um nur die drei N-terminalen und die C-terminalen Aminosäure der Peptide zu bewerten. In Abbildung 19 wird gezeigt, dass diese Art der Affinitätsvorhersage vernünftige Ergebnisse für Peptide mit Längen zwischen 10 und 18 Aminosäuren liefert.

Es ist wichtig, längere Peptide als 9mere zu betrachten, da gezeigt worden ist, dass verschiedene präsentierte MHC Class I Epitope als N-terminal verlängerte Peptidvorläufer in das ER von TAP transportiert werden, wo ERAAP die Vorläufer trimmt, indem es N-terminale Residuen entfernt[63, 64, 65]. So entstehen im ER Peptide mit einer zur Bindung an die MHC Class I Moleküle passenden Länge. Die signifikante Verbesserung der Vorhersage von Epitopen aus Proteinsequenzen, die beim Einbeziehen von längeren Peptiden als Vorläufer des Epitops gefunden worden ist (Abbildung 23), deutet darauf hin,

dass ein großer Teil der Epitope als Vorläufer in das ER transportiert wird. Dieses unterstreicht den Einfluss, den TAP auf die Selektion der MHC Class I Epitope hat.

Da nicht bekannt ist, welches Vorläuferpeptid eines Epitops *in vivo* vom Proteasom hergestellt wird, mittelt die entwickelte Methode über die Beiträge aller möglichen Vorläufer bis zu einer bestimmten Länge. Man erhält so einen verallgemeinerten TAP Transportwert anstelle von Affinitäten. Interessanterweise kann man mit diesem Modell am besten MHC Class I Epitope aus Proteinsequenzen identifizieren, wenn die Bewertung der N-terminalen Aminosäuren im Verhältnis zum C-Terminus schwächer gewichtet wird.

In dem hier erstellten Modell zur Vorhersage des TAP Transports von Peptiden gibt es zwei frei wählbare Parameter: der Gewichtsfaktor α für den N-Terminus des Peptids und die Länge L , bis zu der Vorläufer eines potentiellen 9mer Epitops betrachtet werden. Auf dem HLA-X Datensatz liefern die Werte L gleich 10 und α gleich 0.2 die beste Trennung zwischen Epitopen und Zufallspeptiden und werden daher für die anderen Datensätze verwendet. Diese Parameter variieren jedoch stark, wenn man einzelne Allele des HLA-X Datensatzes betrachtet (Tabelle 9) und werden stark durch die individuelle Bindungspräferenz der einzelnen MHC Class I Allele beeinflusst. Der HLA-X Datensatz beinhaltet viele menschliche Allele und mittelt daher über die Bindungspräferenzen der Allele. So sollten die Parameter, die auf dem Gesamtdatensatz bestimmt worden sind, optimal für ein mittleres MHC Allel sein und den wahren Einfluss des Transports durch TAP exakter wiedergeben. Auf dem H2-X Datensatz sind α gleich 0.02 und L gleich 11 als optimal bestimmt worden. So unterscheiden sich die optimalen Parameter auf den unterschiedlichen Datensätzen recht stark. Wenn man die auf dem HLA-X Datensatz bestimmten Werte verwendet, ist der Verlust an Vorhersagequalität relativ gering ($\Delta AUC < 0.006$) im Vergleich zur Verringerung der Vorhersagequalität, wenn keine Vorläuferpeptide betrachtet werden (L gleich 9) und der N-Terminus nicht im Verhältnis zum C-Terminus gewichtet wird (α gleich 1.0); $\Delta AUC \sim 0.100$. Offenbar sind beide Parameter bedeutungsvoll, jedoch können sie nicht in einem engen Bereich festgelegt werden. Hier wird die Verwendung der auf dem HLA-X Datensatz optimierten Parameter empfohlen, auch wenn L gleich 10 aus einem biologischen Standpunkt als zu gering erscheint, da gezeigt worden ist, dass längere Vorläufer von Epitopen von TAP transportiert werden[129].

TAP transportiert verschiedene Peptide mit sehr unterschiedlichen Transportraten. Abbildung 20 zeigt, dass die TAP Affinität und damit die Transportrate sich maximal um den Faktor $1.45 \cdot 10^7$ differieren kann. TAP kann so eventuelle vom Proteasom hergestellte Mengenverhältnisse zwischen den Peptiden ausgleichen oder verstärken. Es stehen leider keine quantitativen Daten für die Präsentation von Epitopen und aufgrund des schnellen Abbaus[88] keine Konzentrationen von Peptiden im Cytosol zur Verfügung, so dass hier

ein rein qualitativ beschreibender Ansatz gewählt werden muss. Die Bindungsaffinitäten verschiedener Peptide variieren für die MHC Class I Moleküle noch viel stärker als für TAP, so dass Peptide interessant sein können, die vom Proteasom in geringen Mengen erzeugt werden.

Die hier vorgestellten Ergebnisse konnten im Journal of Immunology publiziert werden[87].

3.3 Proteasomale Schnittpräferenz

Cytosolische Proteine werden durch das Proteasom in kurze Peptide gespalten („verdaut“), die dann durch cytosolische Aminopeptidasen zu Aminosäuren abgebaut werden. Das Proteasom erzeugt in diesem Prozess die später präsentierten Antigene beziehungsweise N-terminal verlängerte Precursoren. In diesem Abschnitt der Arbeit soll ein Modell für die Schnittpräferenz des Proteasoms entwickelt werden.

An dieser Stelle soll auf den Unterschied zwischen der Vorhersage von potentiellen Schnittstellen in einem Protein und den erzeugten Fragmenten hingewiesen werden. Die Vorhersage einer Schnittstelle gibt die Wahrscheinlichkeit, mit der das Proteasom an einer bestimmten Peptidbindung im Protein schneidet. Zur Erzeugung eines Peptids durch das Proteasom bedarf es in der Regel zweier Schnitte N- und C-terminal sowie keiner Schnittprozesse im inneren des Peptids. Für die Vorhersage von Schnittstellen kann man und wird hier ein statistisches Modell entwickeln. Für die Vorhersage von Fragmenten bedarf es einer mechanistischen Vorstellung der Prozessierung des abzubauenen Proteins. Untersuchungen zur Generierung von Peptiden werden in einem späteren Kapitel vorgestellt (Kapitel 4).

3.3.1 Verwendete experimentelle Daten

Hier werden die zum Erstellen des Modells verwendeten Daten vorgestellt.

Quantifizierte Verdaus von drei kompletten Proteinen sind der Trainingsdatensatz

Als Trainingsdaten wurden die *in vitro* Verdaus von drei kompletten Proteinen (Enolase[130], β -Casein[131] und Schaf Prion Protein[132]) verwendet. Die beim Verdaus entstandenen Fragmente wurden durch quantitativen Edman-Abbau⁷ identifiziert und quantifiziert. Aus den Mengen der Fragmente wurde die Schnitthäufigkeit an den Enden

⁷ Beim Edman-Abbau wird die N-terminale Aminosäure eines Peptids mit Phenylisothiocyanat (PTH) markiert und anschließend abgespalten, wobei der Rest des Peptids intakt bleibt. Die abgespaltene PTH-Aminosäure kann dann identifiziert und quantifiziert werden.

berechnet und die, die Schnittstelle umgebende Sequenz des Proteins wurde bestimmt. Wenn die berechneten Schnitthäufigkeiten an einer Stelle unterschiedliche Werte hat, je nachdem, ob alle Peptide aufsummiert wurden, deren C-Terminus an dieser Stelle ist, oder alle Peptide mit N-Terminus an dieser Stelle, wurde der höhere Wert genommen.

Beim Verdau der Proteine wurde sowohl das Immunoproteasom als auch das konstitutive Proteasom verwendet. Die Daten wurden dementsprechend unterteilt.

Qualitative Peptidverdaus und HIV nef Proteinverdau sind Testdatensatz

Als Testdatensatz wurden die *in vitro* Verdaus von verschiedenen Peptiden[129, 133, 134, 135, 136, 137, 138, 139], zuvor unveröffentlichten Peptidverdaus[140] und des HIV nef Protein[141] verwendet. Die Schnittstellen beziehungsweise Fragmente wurden hier nur identifiziert aber nicht quantifiziert. Jede Stelle der Sequenz der Peptide wurde als „Schnittstelle“ oder „nicht Schnittstelle“ klassifiziert und dazu die umgebende Sequenz notiert. Die Peptide werden nach der Art des eingesetzten Proteasoms (immuno, konstitutiv) in verschiedene Datensätze unterteilt. Die Datensätze sind in Tabelle 13 zusammengefasst. Weitere Informationen über die verwendeten Peptide ist dem Supplement der Arbeit von Tenzer et al.[140] zu entnehmen (<http://www.mhc-pathway.net/supplement>).

Datensatz von MHC Liganden: SYF-human

Aus der SYFPEITHI Datenbank[106] wurden alle MHC Klasse I Liganden aus menschlichen Proteinen mit einer Länge von 8 bis 12 Aminosäuren extrahiert. Die zugehörigen Proteine wurden mittels Blast identifiziert und die Sequenzen der GenBank Datenbank[142] entnommen. Falls es keine exakte Übereinstimmung des Liganden mit der Proteinsequenz gab oder in der Proteinsequenz unbestimmte Aminosäuren enthalten waren, wurde der MHC Ligand aus dem Datensatz entfernt. Der Datensatz der Liganden und der Proteine wird als SYF-human bezeichnet.

390 zuvor unbekannte MHC Liganden mit bekanntem erzeugendem Proteasom: der RCC Datensatz

Von 3 Patienten mit Nierenzellkarzinom (RCC68, RCC75 und RCC98) wurden 297 neue MHC I Liganden identifiziert. Für RCC75 und RCC98 konnten die meisten der durch Massenspektroskopie identifizierten Liganden anhand von in der Literatur beschriebenen Motiven den von den Patienten exprimierten MHC I Allelen zugeordnet werden. Für RCC68 war das nicht möglich, da die exprimierten MHC I Allele überlappende Bindungsmotive aufwiesen. Der Proteasomsubtyp (immuno oder konstitutiv) ist für diese Zelllinien be-

stimmt worden. Komplettiert wird dieser Datensatz durch 93 weitere MHC I Liganden mit bekanntem Proteasomsubtyp der präsentierenden Zelllinien. Wie auch im SYF-human Datensatz sind auch die Quellproteine der Liganden Bestandteil des Datensatzes. Die Länge der Liganden variiert zwischen 7 und 12 Aminosäuren, wobei hauptsächlich 9mere (70%) im Datensatz enthalten sind. Die Daten wurden entsprechend des Proteasomsubtyps in RCC-c (konstitutiv) und RCC-i unterteilt.

Tabelle 10: RCC Datensatz mit Proteasomsubtyp

Zelllinie	Anzahl	Proteasomsubtyp
RCC68	123	immuno
RCC75	79	konstitutiv
RCC98	95	immuno
LCL	23	immuno
JY, MGAR	70	konstitutiv

RCC ist renal carcinoma cell; LCL, JY und MGAR sind unterschiedliche B-Zell Lymphozyten Zelllinien

3.3.2 Methode der Modellentwicklung

Ausgehend von der Modellannahme, dass die die Schnittstelle umgebenden Aminosäuren unabhängig voneinander die Schnittstärke bestimmen, wurde wieder die SMM Methode verwendet, um ein statistisches Modell zu erstellen.

Die Schnittstärke ergibt sich nach Gleichung 25 aus der die Schnittstelle umgebende Sequenz. Beim Trainieren der Parameter wird der Abstand zwischen den beobachteten Schnittstärken und den vom Modell vorhergesagten (Gleichung 26) minimiert. Für Schnittstellen, an denen experimentell keine Fragmente gefunden wurden, wurden 5 pmol als Nachweisgrenze und damit als experimenteller Wert dieser Schnittstelle festgelegt. Falls weniger als 5 pmol für solch eine Schnittstelle vorhergesagt werden, ist der Abstand zwischen Modell und Experiment gleich null. Der regulierende Parameter λ (Gleichung 27) verringert die Parameter der Matrix und reduziert dadurch Overfitting.

Unterschiedlich zur Verwendung der SMM Methode in vorhergehenden Abschnitten dieses Kapitels ist λ hier Sequenzpositionsabhängig, dass heißt alle Parameter der Matrix, die für eine Position der Sequenz verantwortlich sind werden mit dem gleichen λ reguliert aber für verschiedene Positionen hat λ unterschiedliche Werte. Dieses dient dem Zweck zu bestimmen, wie groß das Sequenzfenster um die Schnittstelle gewählt werden muss. Wenn eine Position keinen signifikanten Beitrag zur Schnittstärke leistet, hat das optimale

λ einen hohen Wert und die zugehörigen Parameter der Bewertungsmatrix haben entsprechend kleine Werte.

Die optimalen Werte für λ_{pos} werden durch 5-fach Kreuzvalidierung auf dem Trainingsdatensatz bestimmt. Ausgehend von optimalen λ_{pos} unter der Annahme, dass die Vorhersage nur von der zugehörigen Position der Sequenz abhängen würde, werden die endgültigen λ_{pos} durch ein Gradientenverfahren optimiert. Dieser Vorgang wurde 20-mal wiederholt, jedes Mal mit unterschiedlicher zufälliger Aufteilung des Trainingsdatensatzes bei der 5-fach Kreuzvalidierung. Die endgültige Vorhersagematrix ist der Mittelwert der 20 unabhängig voneinander bestimmten Matrizen.

Es werden zwei Vorhersagematrizen erstellt. Jeweils eine für das konstitutive und eine für das Immunoproteasom. Für das Training des konstitutiven Proteasommodells wurden die Daten aus dem β -Casein Verdau[131] sowie Daten aus dem Verdau der Enolase[130] mit dem konstitutiven Proteasom verwendet. Für das Immunoproteasommodell wurden ebenfalls die Daten des β -Casein Verdau verwendet, obwohl der Abbau dieses Proteins mit dem konstitutiven Proteasom durchgeführt wurde, sowie die Daten des Abbaus der Enolase mit dem Immunoproteasom. Es wurden nur die Daten aus dem Abbau dieser beiden Proteine verwendet, um zum einen mit anderen Methoden, die die gleichen Daten zum Training verwendet haben vergleichbar zu sein und zum anderen stellt der Prionabbau einen guten Testdatensatz für alle Vorhersagemethoden dar. Die Verwendung der Enolase Daten für das Immunoproteasommodell war notwendig, um ausreichend Daten für das Training der Parameter zu haben.

3.3.3 Ergebnisse

Modelle für das konstitutive und das Immunoproteasom

Signifikante Beiträge zur Schnittstärke kommen aus einem Sequenzfenster von 10 Aminosäuren um die Schnittstelle. Dabei sind 6 Aminosäuren N-terminal von der Schnittstelle (P6 bis P1) und die verbleibenden 4 C-terminal (P1' bis P4').

Tabelle 11: Vorhersagematrix für das konstitutive Proteasom

	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'		P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'		P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	
A	-0.01	0.00	0.07	-0.06	-0.05	0.12	0.26	-0.10	0.00	-0.02	A	-0.04	-0.00	0.04	-0.02	-0.04	0.04	0.37	-0.07	0.04	0.00	A	-0.04	-0.00	0.04	-0.02	-0.04	0.04	0.37	-0.07	0.04	0.00	A
C	-0.00	-0.00	-0.01	-0.01	0.10	-0.04	-0.02	-0.03	-0.00	-0.01	C	-0.04	-0.01	0.02	-0.02	0.03	-0.04	-0.13	-0.05	-0.01	-0.00	C	-0.04	-0.01	0.02	-0.02	0.03	-0.04	-0.13	-0.05	-0.01	-0.00	C
D	0.06	0.00	-0.07	-0.13	-0.13	0.44	-0.10	0.33	-0.00	-0.02	D	0.04	-0.03	-0.12	-0.09	-0.22	0.39	-0.16	0.20	0.00	-0.00	D	0.04	-0.03	-0.12	-0.09	-0.22	0.39	-0.16	0.20	0.00	-0.00	D
E	0.01	-0.00	0.01	0.02	-0.05	0.06	-0.09	-0.04	0.01	-0.06	E	0.03	-0.01	0.00	-0.00	-0.09	-0.14	-0.15	0.01	0.02	-0.02	E	0.03	-0.01	0.00	-0.00	-0.09	-0.14	-0.15	0.01	0.02	-0.02	E
F	-0.04	-0.00	0.04	0.01	0.07	0.06	0.04	0.06	-0.00	0.03	F	-0.13	-0.03	0.00	0.00	0.05	0.10	0.00	0.08	-0.02	0.00	F	-0.13	-0.03	0.00	0.00	0.05	0.10	0.00	0.08	-0.02	0.00	F
G	-0.06	0.00	0.02	-0.12	0.01	-0.19	-0.07	0.01	0.00	-0.04	G	-0.11	0.08	0.21	-0.05	-0.04	-0.30	0.12	0.02	-0.05	-0.00	G	-0.11	0.08	0.21	-0.05	-0.04	-0.30	0.12	0.02	-0.05	-0.00	G
H	0.01	-0.00	-0.02	0.03	0.09	-0.10	-0.02	0.06	-0.00	-0.03	H	0.01	0.03	-0.01	-0.02	0.29	-0.32	-0.07	-0.12	0.01	0.01	H	0.01	0.03	-0.01	-0.02	0.29	-0.32	-0.07	-0.12	0.01	0.01	H
I	0.04	-0.01	0.04	0.01	-0.04	-0.02	-0.10	0.14	-0.00	0.02	I	0.04	-0.06	0.02	0.04	-0.11	-0.10	-0.18	0.04	-0.02	0.00	I	0.04	-0.06	0.02	0.04	-0.11	-0.10	-0.18	0.04	-0.02	0.00	I
K	-0.08	-0.01	-0.07	0.12	-0.00	-0.37	0.00	-0.29	-0.01	0.04	K	-0.05	-0.02	-0.14	0.10	-0.09	-0.53	-0.04	-0.22	-0.05	0.01	K	-0.05	-0.02	-0.14	0.10	-0.09	-0.53	-0.04	-0.22	-0.05	0.01	K
L	-0.02	-0.00	0.01	0.12	-0.12	0.39	-0.06	0.22	0.01	0.00	L	-0.01	-0.02	0.01	0.05	-0.18	0.37	-0.07	0.17	0.03	-0.00	L	-0.01	-0.02	0.01	0.05	-0.18	0.37	-0.07	0.17	0.03	-0.00	L
M	-0.01	-0.00	0.01	-0.07	0.00	0.07	0.10	0.04	0.00	-0.00	M	0.02	-0.01	0.06	-0.05	0.21	0.24	0.11	0.04	-0.01	0.00	M	0.02	-0.01	0.06	-0.05	0.21	0.24	0.11	0.04	-0.01	0.00	M
N	-0.03	-0.00	-0.01	-0.01	0.00	-0.17	-0.01	-0.06	0.00	-0.02	N	-0.06	-0.05	-0.11	0.09	0.09	-0.20	0.08	0.06	-0.01	0.00	N	-0.06	-0.05	-0.11	0.09	0.09	-0.20	0.08	0.06	-0.01	0.00	N
P	0.09	0.01	0.08	-0.02	-0.31	-0.40	-0.03	-0.12	-0.01	0.02	P	0.22	0.11	0.09	0.04	-0.38	-0.78	-0.03	-0.05	0.00	0.00	P	0.22	0.11	0.09	0.04	-0.38	-0.78	-0.03	-0.05	0.00	0.00	P
Q	-0.05	0.01	-0.04	0.01	0.06	-0.16	-0.09	-0.04	0.00	0.05	Q	-0.07	-0.00	-0.09	-0.04	0.25	-0.38	-0.10	0.04	0.02	-0.01	Q	-0.07	-0.00	-0.09	-0.04	0.25	-0.38	-0.10	0.04	0.02	-0.01	Q
R	-0.00	-0.01	-0.03	0.05	0.07	0.02	0.09	-0.07	-0.00	0.01	R	0.09	-0.02	-0.01	0.03	0.11	-0.22	0.11	-0.03	0.04	-0.00	R	0.09	-0.02	-0.01	0.03	0.11	-0.22	0.11	-0.03	0.04	-0.00	R
S	0.05	0.00	-0.05	-0.15	0.01	-0.04	0.04	-0.14	-0.01	0.01	S	-0.02	0.04	-0.03	-0.11	-0.00	-0.25	0.06	-0.09	-0.03	-0.00	S	-0.02	0.04	-0.03	-0.11	-0.00	-0.25	0.06	-0.09	-0.03	-0.00	S
T	0.05	0.00	0.00	0.01	0.15	-0.12	-0.01	0.03	-0.01	0.00	T	0.06	-0.02	-0.05	-0.02	0.11	-0.37	0.00	-0.01	-0.03	-0.00	T	0.06	-0.02	-0.05	-0.02	0.11	-0.37	0.00	-0.01	-0.03	-0.00	T
V	-0.01	0.00	0.00	0.14	0.25	0.12	-0.02	-0.02	0.00	0.01	V	-0.04	-0.02	0.02	0.06	0.17	-0.11	-0.15	-0.06	0.05	0.00	V	-0.04	-0.02	0.02	0.06	0.17	-0.11	-0.15	-0.06	0.05	0.00	V
W	0.01	-0.00	0.00	0.01	-0.06	0.14	0.02	-0.01	-0.00	-0.03	W	0.01	-0.05	0.00	-0.02	-0.10	1.66	0.04	-0.04	-0.00	-0.01	W	0.01	-0.05	0.00	-0.02	-0.10	1.66	0.04	-0.04	-0.00	-0.01	W
Y	0.02	0.00	0.02	0.04	-0.05	0.19	0.06	0.01	0.01	0.03	Y	0.03	0.10	0.09	0.03	-0.04	0.94	0.20	0.09	0.01	0.01	Y	0.03	0.10	0.09	0.03	-0.04	0.94	0.20	0.09	0.01	0.01	Y
Offset:											Offset:											Offset:											
2.16											2.31											2.31											

Auf der linken Seite ist die Vorhersagematrix für das konstitutive Proteasom trainiert auf Enolase- und Casein-Daten dargestellt, rechts die Matrix, trainiert auf Enolase-, Casein- und Prion Protein-Daten. Die Einträge sind analog zum TAP Modell (Tabelle 6) logarithmiert.

Tabelle 12: Vorhersagematrix für das Immunoproteasom

	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'		P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'		P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	
A	-0.04	0.01	0.03	-0.13	0.01	0.14	0.32	0.04	0.05	0.01	A	-0.08	0.01	0.09	-0.07	-0.01	0.08	0.40	0.00	0.17	0.03	A	-0.08	0.01	0.09	-0.07	-0.01	0.08	0.40	0.00	0.17	0.03	A
C	-0.02	-0.00	-0.00	0.06	0.22	0.00	-0.02	-0.09	-0.02	-0.00	C	-0.05	-0.00	-0.09	0.05	-0.05	-0.59	-0.03	-0.10	0.01	-0.02	C	-0.05	-0.00	-0.09	0.05	-0.05	-0.59	-0.03	-0.10	0.01	-0.02	C
D	0.18	0.01	-0.04	-0.12	-0.51	-0.10	-0.15	0.38	-0.01	0.02	D	0.19	-0.01	-0.20	-0.17	-0.39	-0.46	-0.28	0.34	0.01	0.03	D	0.19	-0.01	-0.20	-0.17	-0.39	-0.46	-0.28	0.34	0.01	0.03	D
E	-0.11	-0.00	-0.01	0.04	0.02	-0.32	-0.05	-0.08	0.03	-0.02	E	-0.13	-0.00	-0.09	0.13	0.00	-0.37	-0.21	0.05	0.07	-0.02	E	-0.13	-0.00	-0.09	0.13	0.00	-0.37	-0.21	0.05	0.07	-0.02	E
F	-0.08	-0.00	0.00	-0.11	0.14	0.63	-0.01	-0.02	-0.05	0.01	F	-0.21	-0.00	0.06	-0.11	0.16	0.81	-0.08	-0.09	-0.13	0.02	F	-0.21	-0.00	0.06	-0.11	0.16	0.81	-0.08	-0.09	-0.13	0.02	F
G	0.05	0.00	0.01	-0.23	-0.05	-0.49	0.05	0.01	0.12	-0.00	G	0.06	-0.00	0.10	-0.15	-0.13	-0.66	0.01	-0.14	-0.07	0.01	G	0.06	-0.00	0.10	-0.15	-0.13	-0.66	0.01	-0.14	-0.07	0.01	G
H	0.09	-0.01	0.00	-0.03	0.14	-0.12	-0.02	0.16	0.05	-0.02	H	0.02	0.01	-0.03	-0.12	0.17	-0.25	0.00	-0.19	0.13	0.02	H	0.02	0.01	-0.03	-0.12	0.17	-0.25	0.00	-0.19	0.13	0.02	H
I	0.02	-0.01	0.01	0.16	0.12	0.22	-0.25	0.28	0.02	-0.01	I	0.08	-0.01	0.01	0.24	0.06	0.14	-0.53	0.14	0.02	-0.01	I	0.08	-0.01	0.01	0.24	0.06	0.14	-0.53	0.14	0.02	-0.01	I
K	-0.12	-0.01	-0.02	0.23	0.05	-0.39	0.03	-0.28	-0.10	0.01	K	-0.11	-0.00	-0.10	0.34	-0.01	-0.40	-0.06	-0.14	-0.08	-0.00	K	-0.11	-0.00	-0.10	0.34	-0.01	-0.40	-0.06	-0.14	-0.08	-0.00	K
L	0.09	0.00	-0.00	0.06	-0.13	0.99	-0.29	0.06	0.14	-0.03	L	0.13	0.00	0.01	-0.09	-0.07	0.95	-0.44	0.03	0.14	-0.04	L	0.13	0.00	0.01	-0.09	-0.07	0.95	-0.44	0.03	0.14	-0.04	L
M	-0.05	0.00	0.02	-0.09	0.14	-0.07	0.21	0.04	-0.06	0.01	M	-0.19	-0.00	0.17	-0.02	0.28	0.09	0.46	-0.10	-0.04	-0.00	M	-0.19	-0.00	0.17	-0.02	0.28	0.09	0.46	-0.10	-0.04	-0.00	M
N	-0.01	0.00	-0.00	0.03	0.05	-0.27	0.19	0.02	0.15	0.01	N	0.20	-0.00	0.00	0.21	0.10	-0.44	0.25	0.13	0.13	-0.01	N	0.20	-0.00	0.00	0.21	0.10	-0.44	0.25	0.13	0.13	-0.01	N
P	0.08	0.01	0.01	0.05	-0.79	-0.88	-0.19	-0.07	-0.08	0.00	P	0.24	0.01	0.10	0.12	-0.40	-1.07	0.05	0.04	0.16	0.01	P	0.24	0.01	0.10	0.12	-0.40	-1.07	0.05	0.04	0.16	0.01	P
Q	-0.11	0.00	0.00	0.04	0.10	-0.31	-0.12	-0.04	-0.04	0.00	Q	-0.12	0.00	-0.10	-0.13	0.15	-0.46	-0.03	0.07	0.09	-0.01	Q	-0.12	0.00	-0.10	-0.13	0.15	-0.46	-0.03	0.07	0.09	-0.01	Q
R	0.00	-0.00	-0.00	0.02	0.10	-0.03	0.19	-0.25	-0.07	0.00	R	0.21	-0.00	-0.06	0.03	0.06	-0.01	0.18	-0.11	-0.01	-0.01	R	0.21	-0.00	-0.06	0.03	0.06	-0.01	0.18	-0.11	-0.01	-0.01	R
S	-0.05	-0.00	-0.02	-0.24	0.05	-0.17	0.15	-0.14	-0.08	-0.01	S	-0.15	-0.01	-0.01	-0.27	0.05	-0.35	0.20	-0.01	-0.20	-0.00	S	-0.15	-0.01	-0.01	-0.27	0.05	-0.35	0.20	-0.01	-0.20	-0.00	S
T	0.07	-0.00	-0.01	-0.06	0.31	-0.32	-0.03	0.18	-0.05	-0.01	T	0.20	-0.00	-0.08	-0.18	0.02	-0.42	0.05	0.11	-0.07	-0.02	T	0.20	-0.00	-0.08	-0.18	0.02	-0.42	0.05	0.11	-0.07	-0.02	T
V	-0.05	-0.01	0.01	0.27	0.36	-0.03	-0.06	-0.05	-0.00	0.00	V	-0.11	-0.01	0.08	0.19	0.12	-0.08	-0.34	0.08	0.05	0.01	V	-0.11	-0.01	0.08	0.19	0.12	-0.08	-0.34	0.08	0.05	0.01	V
W	0.05	0.00	0.00	0.01	-0.16	0.84	-0.05	-0.12	-0.05	-0.00	W	-0.12	-0.00	-0.02	-0.17	-0.14	2.26	0.04	-0.17	-0.23	-0.01	W	-0.12	-0.00	-0.02	-0.17	-0.14	2.26	0.04	-0.17	-0.23	-0.01	W
Y	-0.00	0.01	0.00	0.07	-0.17	0.68	0.11	-0.04	0.06	0.01	Y	-0.05	0.00	0.15	0.15	0.03	1.21	-0.04	0.06	-0.14	0.01	Y	-0.05	0.00	0.15	0.15	0.03	1.21	-0.04	0.06	-0.14	0.01	Y
	Offset:					2.39						Offset:					2.58																

ren konnten nicht als signifikant identifiziert werden, sind also an der Vorhersage von Schnittstärken nicht beteiligt.

Vergleich mit anderen Methoden

Zum Zeitpunkt dieser Untersuchungen existierten drei publizierte und öffentlich zugängliche Methoden zur Vorhersage von proteasomalen Schnittstärken basierend auf *in vitro* Daten: FragPredict[143, 144], NetChop 20S[145] und PaProc 1.0[146, 147]. Weiterhin existiert eine Version von PaProc, die zwischen konstitutiven (PaProc 2c) und immuno (PaProc 2i) Proteasom unterscheidet[148]. Alle PaProc Modelle und NetChop 20S wurden mit den Verdaus von β -Casein und Enolase trainiert. FragPredict wurde nachträglich ebenfalls mit diesen Daten trainiert, da die ursprüngliche Version auf dem Abbau einer begrenzten Anzahl von Peptiden basierte.

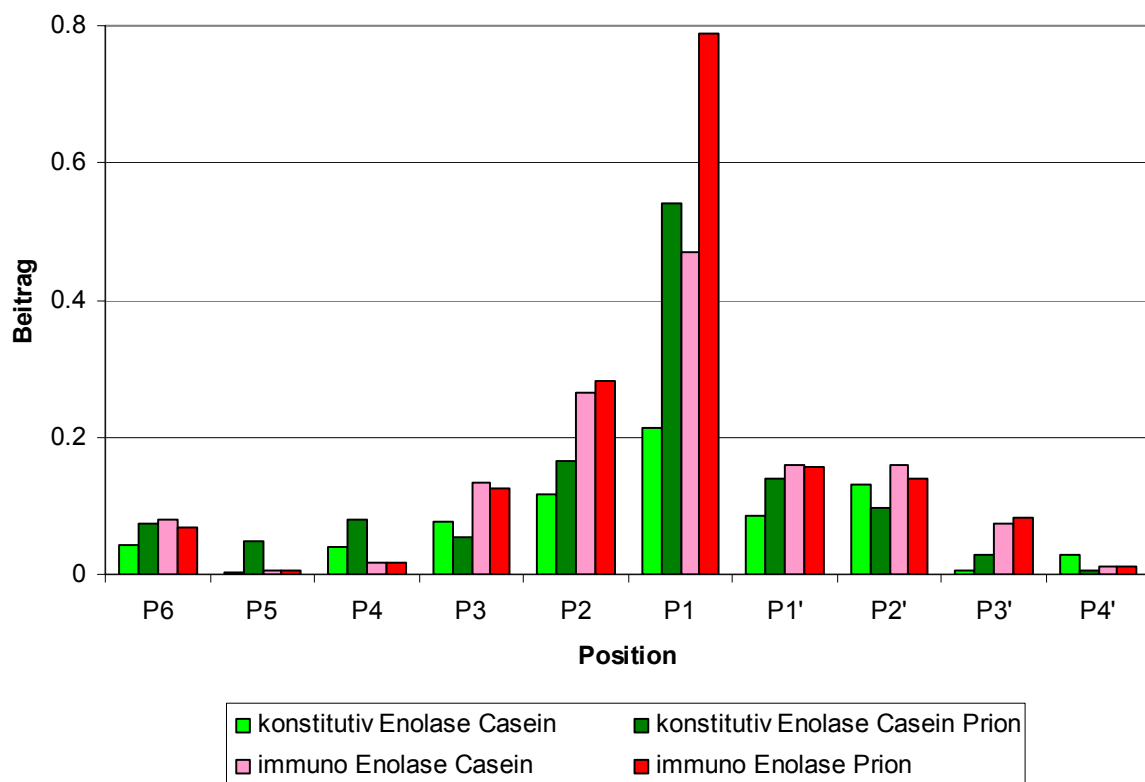


Abbildung 25: Beitrag der Position zur Schnittstärke des Proteasoms

Die Standardabweichung der einzelnen Spalten von Tabelle 11 und Tabelle 12 ist dargestellt. Die Position P1 hat in allen Modellen den größten Beitrag zur Schnittstärke des Proteasoms.

NetChop C2.0 basiert nicht auf *in vitro* Daten, sondern auf der Analyse der C-Termini bekannter MHC Class I Liganden. Auch mit dieser Vorhersagemethode soll die ProteaSMM Methode verglichen werden.

Mit den hier genannten Methoden wurden für die verschiedenen oben erläuterten Testdatensätze Schnittvorhersagen durchgeführt. Abbildung 26 zeigt ein ROC Diagramm der Vorhersagen auf dem gesamten Testdatensatz nur unterteilt in konstitutives und immuno Proteasom. In Tabelle 13 sind die Ergebnisse für die einzelnen Testdatensätze aufgelistet.

Die Analyse dieser Ergebnisse zeigt, dass die hier entwickelten Vorhersagemethoden für das konstitutive Proteasom (ProteaSMM-c) sowie das Immunoproteasom (ProteaSMM-i) besser die experimentellen *in vitro* Daten beschreiben als die bisher etablierten Methoden. Die Verbesserung der Vorhersagequalität von ProteaSMM-c gegenüber NetChop 20S ist nur gering mit einem AUC von 0.71 gegenüber 0.69 von NetChop 20S. Die beiden anderen Methoden sind deutlich schlechter bei der Identifizierung von Schnittstärken. Dieses liegt zum einen daran, dass FragPredict dafür entwickelt wurde Peptidfragmente und nicht Schnittstellen vorherzusagen und zum anderen daran, dass PaProc die Schnittstärke nur in 4 Klassen unterteilt (-, +, ++ und +++) anstatt, wie die anderen Methoden quasi kontinuierliche Schnittstärken zu liefern.

Die Qualität der ProteaSMM-i Methode ist hingegen deutlich besser, als die der anderen Methoden. Überraschend war das sehr gute Abschneiden von ProteaSMM-i auf den Testdaten aus Verdaus mit dem konstitutiven Proteasom mit einem AUC von 0.73. Dieses könnte man damit begründen, dass die verwendeten Daten für das Training des ProteaSMM-i Modells besser für die Modellbildung geeignet sind und auch Daten von konstitutiven Verdaus enthalten. Weiter könnte man spekulieren, dass das Vorhandensein von drei Hauptaktivitäten im konstitutiven Proteasom (tryptisch, chymotryptisch und caspaseähnlich) im Gegensatz zu den zwei Aktivitäten des immuno Proteasoms (tryptisch und chymotryptisch) besser durch das ProteaSMM-i Modell wiedergegeben werden kann, aufgrund der geringeren Komplexität der Trainingsdaten.

Vorhersage von MHC Class I Liganden

Der C-Terminus der meisten MHC I Liganden wird durch das Proteasom hergestellt, während Aminopeptidase den N-Terminus noch trimmen können. Hier wird nun untersucht, wie gut man aus der Schnittstärke des Proteasoms am C-Terminus der Peptide MHC I Liganden vorhersagen kann.

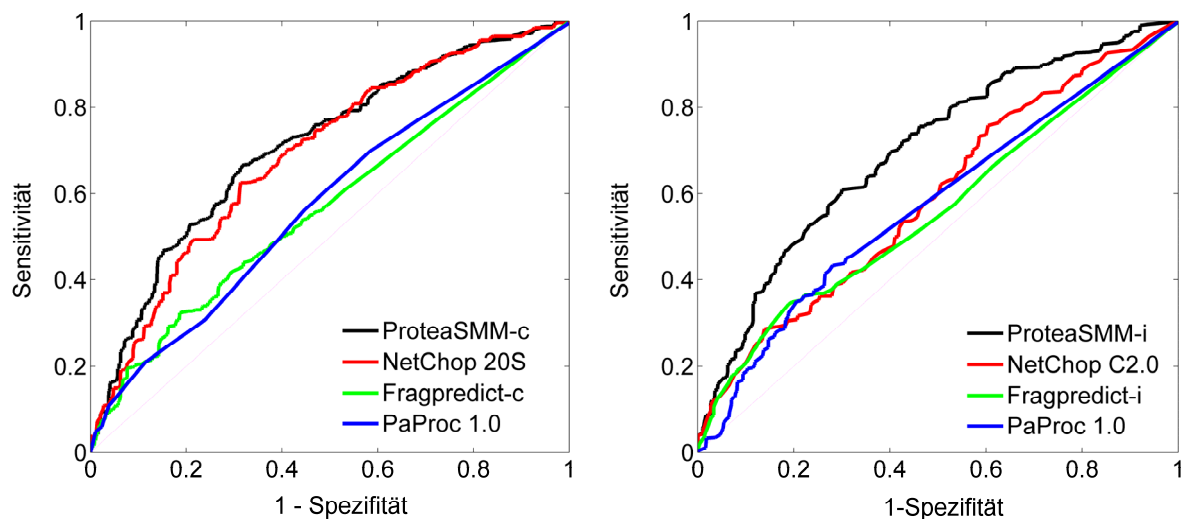


Abbildung 26: Vorhersagequalität für *in vitro* Proteasomverdaus mit verschiedenen Methoden

ROC Kurven der verschiedenen Vorhersagemethoden für die proteasomale Schnitt-rate. Die gestrichelte Linie ist eine Zufallsvorhersage. Die linke Kurve ist die Vorhersage für die Daten des konstitutiven Proteasoms, die rechte für das immuno Proteasom. Die ProteaSMM Methoden haben jeweils die höchste Vorhersagequalität.

Tabelle 13: Vergleich der Güte von Vorhersagemethoden für proteasomale Verdaus *in vitro*

Verdautes Substrat	Verwendeter Proteasomtyp		Vorhersagemethode								
			NetChop		FragPredict		PaProc			ProteaSMM	
	c20S	i20S	20S	C2.0	-c	-i	1	2c	2i	-c	-i
Peptidmix	X		0.71	0.65	0.54	0.56	0.63	0.61	0.60	0.72	0.74
Peptidmix		X	0.74	0.67	0.55	0.63	0.62	0.56	0.58	0.74	0.80
Peptide ssx2	X		0.78	0.70	0.58	0.69	0.56	0.60	0.67	0.81	0.81
Peptide PRAME		X	0.79	0.67	0.68	0.65	0.61	0.57	0.56	0.80	0.82
Prion	X		0.59	0.57	0.60	0.60	0.52	0.60	0.54	0.64	0.68
Prion		X	0.60	0.50	0.55	0.51	0.53	0.63	0.52	0.63	0.65
HIV Nef	X	X	0.63	0.62	0.60	0.61	0.52	0.53	0.57	0.67	0.69
Vorhersage von MHC I Liganden C-Termini			0.69	(0.83)*	0.67	0.68	0.60	0.64	0.58	0.67	0.76

Die Tabelle gibt die AUC Werte der verschiedenen Methoden auf den einzelnen Datensätzen an. Die beste Methode ist Fett hervorgehoben. Bei der Vorhersage von C-Termini von MHC I Liganden ist anzumerken, dass NetChop C2.0 auf solchen Daten trainiert worden ist und somit dieses keinen unabhängigen Test darstellt.

Hierzu wird für die Datensätze SYF-human, RCC-c und RCC-i die Schnittstärke an allen Positionen der enthaltenen Proteine bestimmt. Aus diesen Schnittstärken wird anhand eines Cutoffs entschieden, ob eine Schnittstelle zu einem MHC Liganden führt oder nicht. Durch Variation des Cutoff erhält man die schon bekannte ROC Kurve und den zugehörigen AUC Wert. Die unterste Zeile in Tabelle 13 enthält die AUC Werte der einzelnen Vorhersagemethoden für den Gesamtdatensatz der Liganden. Abgesehen von NetChop C2.0 ist ProteaSMM-i am besten in der Lage zwischen Liganden und Zufallspeptiden zu unter-

scheiden. NetChop C2.0 wurde im Gegensatz zu den anderen Methoden auf den Motiven von MHC Liganden trainiert, wodurch dies kein fairer Vergleich zwischen den Methoden ist.

Die ROC Kurven der ProteaSMM Methoden sind in Abbildung 34 dargestellt und Tabelle 18 zeigt das Abschneiden dieser Methoden auf den einzelnen Teilen des Ligandendatensatzes.

Alle Vorhersagemethoden liefern deutlich höhere AUC Werte als 0.5, dass heißt MHC Liganden oder deren N-Terminus verlängerte Vorläufer werden effizienter durch das Proteasom hergestellt als andere Peptide.

MHC Liganden mit einem C-terminalen Lysin werden schlecht durch das Proteasom erzeugt

Der P1 Rest, der den C-Terminus eines proteolytischen Fragments bildet, bestimmt am stärksten die Schnittstärke des Proteasoms und auch den TAP Transport. In Tabelle 14 ist dargestellt, welche Aminosäure wie häufig am C-Terminus eines MHC Liganden auftritt und welche Proteasom und TAP Bewertungen diese Aminosäuren erhalten. C-Termini, die gut durch das Proteasom geschnitten werden und nicht schlecht durch TAP transportiert werden, haben eine hohe Häufigkeit bei MHC Liganden. Auffällig ist jedoch, dass Liganden mit einem Lysin am C-Terminus (P1K Liganden) eine hohe Häufigkeit besitzen (9.3% aller Liganden), obwohl sie nur im geringen Maße durch das Proteasom hergestellt werden. Der Transport dieser Peptide mittels TAP ist weder günstig noch ungünstig. Zusammengefasst weist dies auf eine mögliche andere Protease hin, die den C-Terminus von P1K Liganden generiert oder auf einen Mechanismus, der dafür sorgt, dass P1K Liganden besonders effektiv prozessiert werden.

Tabelle 14 Vergleich der Proteasom und TAP Scores der P1 Position von MHC Liganden

Amino- säure	ProteaSMM-i Score	ProteaSMM-c Score	Mittlerer Proteasom Score	TAP Score	Anzahl der Epitope
L	0.99	0.39	0.69	0.94	262
W	0.84	0.14	0.49	0.87	20
Y	0.68	0.19	0.43	2.90	174
F	0.63	0.06	0.35	2.51	96
D	-0.10	0.44	0.17	-1.84	2
A	0.14	0.12	0.13	-0.55	22
I	0.22	-0.02	0.10	0.51	54
V	-0.03	0.12	0.04	0.30	118
M	-0.07	0.07	0.00	0.30	20
R	-0.03	0.02	-0.01	1.47	46
C	0.00	-0.04	-0.02	0.00	2
S	-0.17	-0.04	-0.10	-2.26	3
H	-0.12	-0.10	-0.11	-0.55	4
E	-0.32	0.06	-0.13	-1.57	2
N	-0.27	-0.17	-0.22	-1.34	0
T	-0.32	-0.12	-0.22	-0.71	3
Q	-0.31	-0.16	-0.23	-0.12	4
G	-0.49	-0.19	-0.34	-1.40	1
K	-0.39	-0.37	-0.38	0.46	86
P	-0.88	-0.40	-0.64	0.09	6

Die Aminosäuren sind nach dem mittleren Proteasom Score sortiert. Hervorgehoben sind Scores, die geringer als der Durchschnitt der Methode sind. Lysin (K) hat eine höhere Häufigkeit am C-Terminus von MHC I Liganden als aus der Schnitthäufigkeit zu erwarten wäre.

Eine weitere Untersuchung der P1' Position der P1K Liganden zeigt eine deutlich höhere Frequenz von V, K, L und M an dieser Stelle für P1K Liganden als für alle anderen Liganden. Die Aminosäuren E, G, N, Q, S und T haben hingegen geringere Häufigkeiten für P1K Liganden (Abbildung 27). Wenn man in Betracht zieht, dass nur die erzeugende Protease die P1' Aminosäure sieht, weisen diese Ergebnisse auf eine Beteiligung einer anderen Protease als dem Proteasom an der Erzeugung der P1K Liganden hin. Die Untersuchung der P2' und P3' Aminosäurehäufigkeiten von P1K Liganden zeigt ebenfalls Unterschiede zu anderen MHC I Liganden, obwohl diese an diesen Positionen nicht mehr so groß sind. Somit findet sich ein recht guter Hinweis auf eine andere Protease und die Hypothese der besonderen Prozessierung der P1K Liganden kann verworfen werden.

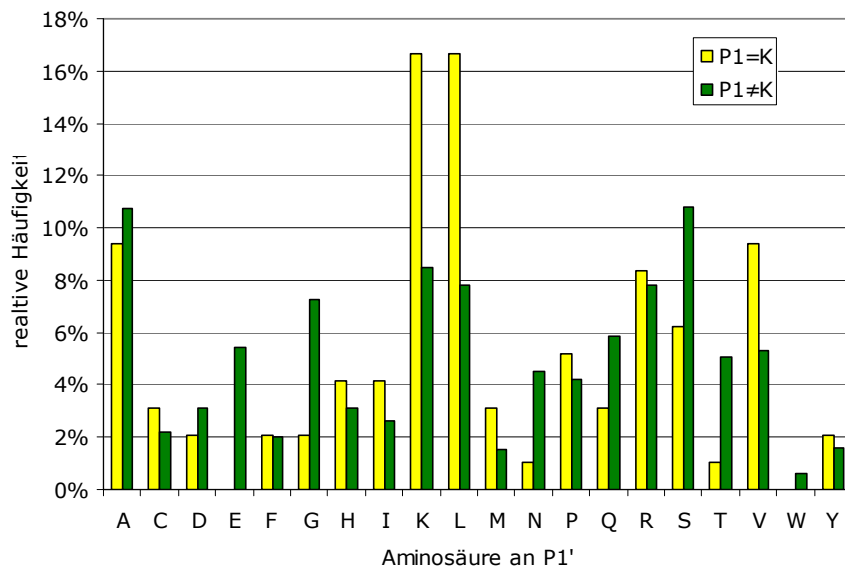


Abbildung 27: Häufigkeit der Aminosäuren an der P1' Position von P1K Liganden

Ein χ^2 Test Analyse zeigt, dass der Unterschied der Verteilungen an der P1' Position für P1K Liganden und allen anderen Liganden signifikant ist ($p=0.0087$). Signifikante Abweichungen ($p<0.05$) gibt es ansonsten nur für die P1' Verteilung von Liganden mit einem Glycin ($p=0.001$) oder Cystein ($p=0.043$) am C-Terminus, jedoch sind hier die Anzahlen der Liganden (Tabelle 14) zu gering, um verlässliche Aussagen zu treffen.

3.3.4 Diskussion

Das neue Schnittstellenvorhersagemodell für das Proteasom zeigt die Selektivität des ersten Schrittes der Antigenpräsentation

Der C-Terminus der meisten MHC I Liganden wird durch das Proteasom hergestellt und die proteasomale Erzeugung der Liganden oder N-terminal verlängerter Precursoren ist der erste Schritt der Antigenpräsentation im MHC Class I Pathway[149]. Hier wurde ein Algorithmus entwickelt, der es erlaubt, proteasomale Schnittraten sowohl für das konstitutive als auch das immuno Proteasom vorherzusagen und zu analysieren.

Das erstellte Modell hat eine höhere Vorhersagequalität als andere etablierte Modelle bei der Vorhersage von *in vitro* Verdaus von kleinen Peptiden, ganzen Proteinen und der Vorhersage von C-Termini von MHC Liganden. Der letzte Punkt ist besonders interessant, da das schlechte Abschneiden der zuvor etablierten Methoden die Frage aufwarf, ob das Proteasom überhaupt selektiv für die Antigenpräsentation ist[87]. Die hier durchgeführte Analyse zeigt, dass das Proteasom durchaus nur eine Auswahl an Peptiden erzeugt, die

dann für die Antigenpräsentation zur Verfügung steht aber dieser Schritt nicht so restriktiv ist wie der Transport mittels TAP oder gar MHC Molekül Bindung. So unterscheiden sich die zwei Schnittraten maximal um einen Faktor von $1.75 \cdot 10^3$ für das immuno Proteasom im Gegensatz zu einem Faktor von $1.45 \cdot 10^7$ für den Transport zweier Peptide mittels TAP oder $2.41 \cdot 10^{15}$ für die Bindung an ein HLA-0201 MHC Molekül.

Dadurch, dass dieses Modell ausschließlich mit in vitro Daten von proteasomalen Verdaus trainiert worden ist, beschreibt es auch nur das Verhalten des Proteasoms in der Antigenpräsentation und nicht andere involvierte Prozesse. Dieses und die Tatsache, dass jeder potentiellen Schnittstelle eine Schnittrate zugewiesen werden kann, macht das Modell zu einem wertvollen Modul für die Modellierung des MHC Class I Pathways.

Tripeptidyl Protease II (TPPII) ist potentielle Quelle von antigenen Peptiden

Die beobachtete Übereinstimmung der hier gemachten Vorhersagen und der Verteilung der MHC I Liganden, insbesondere ihrer C-Termini, zeigt, dass entweder der Einfluss von alternativen Wegen zur Antigenpräsentation[67] gering ist oder die Selektivität der nicht klassischen Wege der des Proteasom und TAP Systems ähnlich ist. Eine Alternative für die Erzeugung der C-Termini von MHC Liganden beinhaltet TPPII[69, 150]. Diese Protease scheint hauptsächlich für das N-terminale Trimmen von längeren Peptiden im Cytosol, den primären proteasomalen Produkten[151], zuständig zu sein. Allerdings ist diese Protease auch in der Lage, den C-Terminus eines MHC Liganden zu generieren, wenn das N-terminale Trimmingprodukt lang genug ist, um ein MHC I Epitop zu beinhalten. Für den MHC Liganden HIV-Nef(73-82) wurde gezeigt, dass er exklusiv durch TPPII hergestellt wird[152]. Dieser Ligand enthält ein Lysin am C-Terminus und unsere Analyse deutet darauf hin, dass mehr dieser P1K Liganden von TPPII hergestellt werden und nicht vom Proteasom. Dieses wird dadurch gestützt, dass der Anteil der P1K Liganden an den bekannten MHC Liganden unerwartet hoch ist, da ein Lysin an der P1 Stelle zu einer geringen proteasomalen Schnittrate führt. Außerdem ist die Verteilung von Aminosäuren an den P1', P2' und P3' Positionen von P1K Liganden anders, als für andere MHC Liganden. So sind für andere MHC Liganden die hydrophoben, verzweigtkettigen Aminosäuren Valin, Leucin und Isoleucin kaum an der Position P1' zu beobachten, während sie bei P1K Liganden verstärkt auftreten. Dieses macht es wahrscheinlich, dass eine alternative Protease für die Generierung von P1K Liganden zuständig ist und TPPII ist ein Kandidat mit einer hohen Wahrscheinlichkeit.

Das Immunoproteasom produziert potentielle MHC Liganden mit einer höheren Effizienz als das konstitutive Proteasom

Ein Ziel dieser Arbeit war es zu untersuchen, ob der Subtyp des Proteasoms einen Einfluss auf die Antigenpräsentation hat. Einige *in vitro* Verdaus zeigten unterschiedliche Schnittmuster für das konstitutive und das immuno Proteasom auf [130, 132], welche sich aber in großen Teilen ähnlich sind. Wenn man den Einfluss des Proteasomsubtyps auf die Antigenpräsentation untersucht, wird dieses dadurch erschwert, dass man nicht weiß, welcher MHC I Ligand von welchem Proteasomtyp generiert worden ist. Daher wurde im Zusammenhang mit dieser Arbeit von experimentellen Partnern ein neuer Datensatz von zuvor unbekannten MHC Liganden aufgebaut, für die der Subtyp des erzeugenden Proteasoms bekannt ist. Überraschenderweise ist die Vorhersage von MHC I Liganden immer erfolgreicher, wenn das Modell des immuno Proteasoms verwendet wird, sogar auf dem Datensatz, der vom konstitutiven Proteasom erzeugten MHC Liganden. Den Grund dafür findet man, wenn man die Antigenvorhersage von TAP und Proteasom kombiniert (Abschnitt 3.5.3). Die TAP Vorhersage glättet die Unterschiede zwischen den Schnittmustern der beiden Proteasommodelle. So zeigt Tabelle 18, dass für die Antigenvorhersage ProteaSMM-i mit einem AUC von 0.761 deutlich besser ist, als ProteaSMM-c mit 0.674. In Kombination mit der Vorhersage von TAP verschwindet dieser Unterschied jedoch (AUC Werte immuno 0.831 und konstitutiv 0.827).

Eine Erklärung hierfür ist, dass das konstitutive Proteasom relativ unspezifisch schneidet und dabei eine Reihe von Peptiden erzeugt, die schlecht von TAP transportiert werden können. Insbesondere werden Peptide mit negativ geladenen Aminosäuren (Aspartat und Glutamat) am C-Terminus erzeugt. Solche Peptide werden bei der Kombination der Vorhersagemethoden effektiv herausgefiltert. Das immuno Proteasom erzeugt ein eingeschränktes Spektrum an Peptiden, die jedoch besser von TAP transportiert werden können und so der Antigenpräsentation dienen. Gestützt wird dies durch die Beobachtung, dass die Spezifitäten von TAP am C-Terminus des zu transportierenden Peptids und die P1 Position des immuno Proteasoms deutlich stärker korrelieren ($r^2=0.37$) als TAP und das konstitutive Proteasom ($r^2=0.02$). Die Korrelation zwischen den Spezifitäten der P1 Positionen der beiden Proteasommodelle ($r^2=0.51$) zeigt dabei, dass die Schnittmuster zum großen Teil ähnlich sind.

Diese Überlegungen führen zu der Feststellung, dass das immuno Proteasom nicht im Laufe der Evolution entwickelt wurde, um das Muster der Antigene an der Zelloberfläche zu ändern, sondern um potentielle MHC Class I Liganden mit einer höheren Effizienz zu erzeugen.

Die hier gezeigten Ergebnisse wurden in Cellular and Molecular Life Sciences veröffentlicht[140].

3.4 N-terminales Trimmen von Peptiden

Vom Proteasom erzeugte Peptide werden im Cytosol durch Aminopeptidasen schrittweise hydrolysiert oder mittels TAP in das ER transportiert, wo ebenfalls Aminopeptidasen den N-Terminus eines Peptides durch Abspaltung von Aminosäuren verändern können. Hier soll der Einfluss der cytosolischen und endoplasmatischen Aminopeptidasen auf die Antigenpräsentation sowie ihre Spezifität untersucht werden.

3.4.1 Verwendete Daten und Methoden

MHC Class I Liganden Datensatz: SYFPEITHI-Liganden

Aus der SYFPEITHI Datenbank[106] wurden 1565 menschliche MHC I Liganden zwischen 8 und 11 Aminosäuren Länge sowie die zugehörigen Proteine entnommen. Um Randeffekte zu vermeiden, wurden nur Liganden verwendet, die im Protein an beiden Seiten wenigstens 20 weitere Aminosäuren haben. Dadurch verringert sich der Datensatz der SYFPEITHI-Liganden auf $n=1543$.

Spezifität der Aminopeptidasen

Von experimentell arbeitenden Partnern wurde im Zusammenhang mit dieser Studie die Aktivität der Aminopeptidasen in Cytosol und im ER bestimmt. Hierzu wurden cytosolische und endoplasmatische Fraktionen von sieben Zelllinien gebildet. Die Aktivität wurde gemessen, indem die Fluoreszenz in Einheiten von RFUs⁸ nach Zugabe von X-amino-4-methyl Cumarin verfolgt wurde. X steht hierbei für jede einzelne der 20 proteinogenen Aminosäuren.

Aus diesen experimentellen Rohdaten wurden dann die initialen Umsatzraten der Aminopeptidasen extrahiert. Die Anfangsgeschwindigkeit der Reaktion v_0 wurde berechnet, indem aus den Zeitverläufen der Fluoreszenz der Anstieg des anfänglichen linearen Teils bestimmt wurde. Diese Raten [$\text{RFU} \cdot \text{min}^{-1}$] wurden zuerst auf das verwendete Volumen normiert ($[\text{RFU} \cdot \text{min}^{-1} \cdot \mu\text{l}^{-1}]$) und anschließend auf die gesamte Hydrolyserate in der Fraktion normiert (Gleichung 16).

⁸ RFU: real fluorescence unit

(16)

$$r_{a_i} = \frac{v_{0,a_i}}{\sum_j v_{0,a_j}}$$

Die so erhaltenen prozentualen Anteile r_a der Hydrolysegeschwindigkeit einer n-terminalen Aminosäure an der Gesamthydrolysegeschwindigkeit, machen es möglich, die Daten aus verschiedenen Experimenten und den beiden Fraktionen miteinander zu vergleichen.

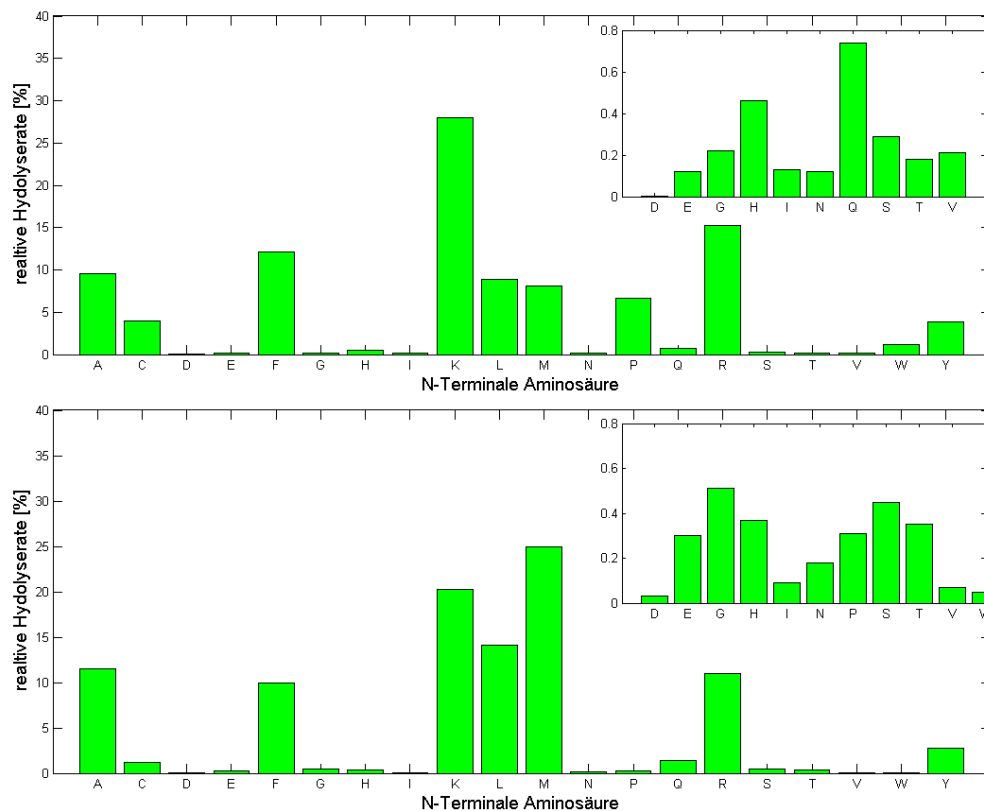


Abbildung 28: Relative Hydrolyseraten der cytosolischen und endoplasmatischen Aminopeptidasen

Durchschnittliche Umsatzraten für alle 20 proteinogenen Aminosäuren (oben: cytosolisch; unten: endoplasmatisch). Aminosäuren mit geringen Schnittraten sind als Inset angegeben. Die Summe aller Schnittraten in einem Kompartiment ist 1.

Bewertungsmatrizen und statistische Maße

Cytosolische Aminopeptidasen modifizieren die N-terminale Sequenz der von TAP transportierten Peptide. Zum Bewerten der Transporteffizienz eines Peptids wird die in Tabelle 6 gezeigte Matrix verwendet. Da hier nur der N-Terminus von Interesse ist, gehen nur die 3 n-terminalen Aminosäuren in die Bewertung ein.

Die Schnittwahrscheinlichkeit, mit der ein bestimmter N-Terminus eines Peptids vom Proteasom erzeugt wird, wird mit den in Tabelle 11 b) und Tabelle 12 b) gezeigten Matrizen berechnet.

Für die Bewertung der Effizienz der Aminopeptidasen werden die Werte, die in Abbildung 28 dargestellt sind, verwendet. Diese Werte werden logarithmiert, um sie mit den TAP Transportraten und den proteasomalen Schnittraten kompatibel zu machen.

Der Ausdruck $f_{\text{Prozess}}(d)_{\text{Ligand}}$ gibt die Bewertung für die Sequenzposition d in Relation zum N-Terminus des Liganden für einen der betrachteten Prozesse. TAP bezeichnet TAP vermittelten Transport der Peptide in das ER, Proteasom steht für das proteasomale Schneiden, wobei Positionen d die P1' Aminosäure des Schnitts ist. Der Prozess des N-terminalen Trimmens durch Aminopeptidasen wird durch AP gekennzeichnet. Die den Liganden flankierenden Sequenzen der enthaltenden Proteine werden verwendet, um die Bewertung durchzuführen. Diese Scores f werden über alle Liganden des SYFPEITHI Datensatzes gemittelt (S), Standardabweichungen (SD) und Standardfehler des Mittelwertes (SEM) berechnet (Gleichung 17).

(17)

$$S_{\text{Prozess}}(d) = \frac{\sum_{i=1}^n f_{\text{Prozess}}(d)_{\text{Ligand}_i}}{n}$$
$$SD_{\text{Prozess}}(d) = \sqrt{\frac{\sum_{i=1}^n (S_{\text{Prozess}}(d) - f_{\text{Prozess}}(d)_{\text{Ligand}_i})^2}{n-1}}$$
$$SEM_{\text{Prozess}}(d) = \frac{SD_{\text{Prozess}}(d)}{\sqrt{n}}$$

Weiterhin werden für alle Liganden an 1000 zufällig gewählten Positionen im Protein Scores berechnet, und gemittelt S^* , um sie mit den Werten an festen Positionen d vergleichen zu können. Auch die Standardabweichung der Zufallsscores SD^* wird berechnet.

3.4.2 Ergebnisse

Die acht Aminosäuren N-terminal von MHC Liganden bilden ein N-terminales Processing Motiv

Es wurden die Häufigkeiten von Aminosäuren bis zu 15 Residuen N-terminal der SYFPEITHI-Liganden untersucht und mit der Häufigkeit der Aminosäuren in allen Quellproteinen verglichen (Tabelle 15). Eine χ^2 -Test Analyse zeigt, dass die Positionen -5 bis -1 vom N-Terminus aus signifikant von der generellen Verteilung abweichen ($p_{19} \leq 0.001$). Solche signifikanten Abweichungen wurden in einer früheren Studie[153] nicht beobachtet, was möglicherweise an der damals geringen Menge an zur Verfügung stehenden Daten lag.

Ein Korrelationsvergleich zwischen den beobachteten Aminosäurehäufigkeiten an verschiedenen Positionen (Tabelle 16) zeigt, dass benachbarte Positionen mit d zwischen -1 und -8 ein ähnliches Profil von Frequenzabweichungen von der Zufallszusammensetzung haben. Dieses schließt auch die Positionen -6 bis -8 mit ein, bei denen die Aminosäurehäufigkeiten nicht signifikant von der Zufallszusammensetzung abweichen. Weiter voneinander entfernte Positionen haben im Gegensatz dazu geringe oder sogar negative Korrelationen.

Diese Analyse zeigt, dass ein Sequenzfenster von etwa 8 Aminosäuren N-terminal von MHC Liganden deutlich eine andere Aminosäureverteilung aufweist, als der Rest der Proteine. Dieses Sequenzfenster wird als „N-terminales Processing Motiv“ bezeichnet. Dieses Motiv spiegelt vermutlich die Präferenzen verschiedener am Prozessieren der Antigene beteiligten Prozesse wieder, wobei diese Präferenzen ähnlich für benachbarte Positionen in der Sequenz sind aber sich mit zunehmender Entfernung ändern.

Das N-terminale Processing Motiv stimmt mit den Spezifitäten von TAP und Proteasom überein

Um zu Analysieren, wie gut sich das N-terminal Processing Motiv durch die Präferenzen von Proteasom und TAP erklären lassen, wurden diese Präferenzen wie in Abschnitt 3.4.1 beschrieben für Precursoren von MHC Liganden berechnet und mit der mittleren Präferenz verglichen (Abbildung 29). Für die Positionen N-terminal vom Liganden gelegene Region bis zu 6 Aminosäuren zeigen sowohl das konstitutive als auch das immuno Proteasom eine erhöhte Schnittrate mit mehr als zwei SEM im Vergleich zum Zufallspeptid. Für TAP werden alle Precursoren des Liganden mit bis zu 8 zusätzlichen Aminosäuren besser in das ER transportiert als zufällige Peptide aus den gleichen Proteinen. Ausnahme ist hier der um zwei Aminosäuren verlängerte Ligand, der keine signifikante Erhöhung

der Transportrate erfährt. Auch für TAP gilt eine Änderung der Rate erst als signifikant, wenn der Mittelwert der Rate für die Position S(d) um mehr als zwei Standardfehler des Mittelwerts SEM(d) von dem Mittelwert für Zufallspeptide S^* abweicht ($p=0.05$).

Das heißt also, dass Aminosäuren, die distal vom N-Terminus von MHC Liganden gefunden werden, das Schneiden durch das Proteasom begünstigen und auch den Transport durch TAP. Somit kann das gefundene N-terminale Processing Motiv von MHC Liganden zumindest teilweise durch die Spezifität des Proteasoms und von TAP erklärt werden.

Tabelle 15: Aminosäurefrequenz N-terminal von MHC I Liganden

	Erwartet	Entfernung zum Epitop N-Terminus															N-Term
		-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	
Aminosäurefrequenz	A	7.3%	7.8%	7.2%	6.8%	6.8%	6.8%	7.1%	6.9%	6.9%	7.1%	7.6%	6.9%	7.1%	7.8%	9.3%	7.5%
	C	1.7%	2.6%	1.5%	0.9%	1.1%	2.1%	1.7%	1.7%	2.1%	1.7%	2.1%	1.6%	1.7%	1.9%	2.5%	0.1%
	D	5.3%	5.4%	6.7%	5.1%	4.6%	4.4%	5.4%	6.2%	5.0%	5.1%	4.5%	4.0%	4.7%	5.0%	6.1%	6.9%
	E	7.8%	6.2%	8.0%	6.2%	6.2%	7.7%	8.4%	7.5%	6.7%	7.3%	7.1%	6.9%	7.5%	8.9%	8.5%	7.7%
	F	3.5%	2.6%	3.5%	4.1%	3.8%	3.3%	4.1%	3.9%	4.1%	3.9%	4.8%	4.5%	3.2%	2.7%	4.0%	3.4%
	G	6.3%	6.0%	7.3%	6.3%	7.8%	6.4%	6.9%	7.4%	6.9%	6.5%	6.0%	5.6%	5.2%	4.7%	6.4%	5.7%
	H	2.4%	2.8%	2.0%	2.3%	3.1%	2.9%	2.0%	2.8%	2.6%	2.8%	2.3%	1.9%	2.9%	2.7%	1.4%	1.9%
	I	4.7%	5.0%	5.6%	4.7%	3.7%	4.7%	4.9%	4.9%	6.4%	5.4%	5.7%	6.0%	6.4%	4.7%	3.0%	3.8%
	K	6.4%	7.6%	6.1%	5.7%	7.8%	6.7%	6.7%	5.5%	6.0%	6.4%	5.9%	6.2%	6.0%	7.4%	4.6%	6.5%
	L	9.8%	10.3%	9.2%	11.3%	10.4%	9.4%	8.2%	11.0%	9.2%	10.7%	10.8%	12.2%	10.8%	9.7%	10.0%	13.0%
	M	2.4%	2.3%	2.2%	2.2%	3.0%	1.8%	2.3%	2.7%	2.5%	2.4%	2.9%	1.9%	2.7%	3.4%	3.0%	1.4%
	N	3.7%	3.6%	4.6%	2.8%	4.0%	4.7%	4.0%	3.3%	3.8%	3.2%	2.9%	3.8%	2.7%	4.1%	4.3%	4.3%
	P	5.5%	4.8%	4.0%	5.0%	5.6%	5.4%	6.1%	5.4%	4.7%	5.4%	5.4%	4.0%	4.1%	3.2%	2.2%	1.7%
	Q	4.9%	4.4%	4.7%	4.6%	4.2%	4.9%	4.3%	5.1%	5.3%	3.9%	4.1%	5.2%	5.1%	4.7%	4.2%	2.5%
	R	5.5%	5.4%	5.1%	6.6%	6.3%	4.3%	4.6%	5.2%	5.5%	6.0%	5.2%	6.4%	5.2%	6.9%	6.0%	5.6%
	S	7.5%	7.1%	7.5%	6.8%	6.2%	6.6%	6.2%	5.7%	5.3%	6.4%	6.5%	5.9%	6.2%	6.8%	7.4%	9.2%
	T	5.2%	4.2%	5.2%	5.4%	4.6%	5.6%	7.3%	3.8%	6.4%	4.8%	5.6%	4.7%	6.0%	6.0%	5.4%	3.8%
	V	6.3%	7.5%	6.1%	8.1%	6.7%	8.0%	6.5%	7.0%	5.7%	7.0%	7.2%	7.4%	6.5%	6.6%	5.7%	4.9%
	W	1.1%	1.0%	0.9%	1.2%	0.8%	0.7%	0.8%	1.0%	1.6%	1.6%	1.6%	1.8%	2.0%	1.4%	0.8%	1.0%
	Y	2.8%	3.4%	2.7%	3.9%	3.0%	3.6%	2.8%	2.5%	3.8%	2.5%	3.3%	4.3%	3.1%	4.0%	3.3%	5.0%
p-Wert		1.8×10^{-2}	2.0×10^{-1}	1.2×10^{-3}	2.0×10^{-3}	1.8×10^{-2}	3.7×10^{-2}	4.1×10^{-2}	2.9×10^{-3}	3.1×10^{-1}	6.4×10^{-2}	4.9×10^{-6}	1.2×10^{-4}	2.7×10^{-5}	1.8×10^{-16}	3.2×10^{-31}	0.00
ChiWert		35.6	23.4	41.6	40.2	33.8	32.6	27.8	41.8	22.2	31.3	60.8	59.1	46.1	68.7	115.0	344.7

Die beobachteten Häufigkeiten der 20 proteinogenen Aminosäuren (F_{pos}) N-terminal von den 1543 bekannten MHC I Liganden sind in der Tabelle angegeben. Die Häufigkeit der Aminosäuren in den zugrund liegenden Proteinen ist in der ersten Spalte aufgetragen (F_{exp}). Häufigkeiten, die um mehr als zwei Standardabweichung von der erwarteten Häufigkeit Abweichen, sind farblich markiert (grün: erhöht; rot: verringert). Die untersten zwei Zeilen geben den p-Wert und den χ^2 -Wert des χ^2 -Tests für die Verteilung der Häufigkeiten an dieser Position im Vergleich zur Verteilung der Häufigkeiten der Aminosäuren über die gesamten Proteine an. Sequenzpositionen gelten als signifikant unterschiedlich, wenn der p-Wert kleiner als 0.001 ist.

Prozessieren von Peptiden durch das Proteasom und durch TAP allein reicht nicht aus, um die beobachtete N-terminale Verteilung von Aminosäuren zu erklären

Es ist unwahrscheinlich, dass das N-terminale Processing Motiv ausschließlich von den Präferenzen des proteasomalen Proteinabbaus und des Transports mittels TAP abhängt. Wenn N-terminal verlängerte Precursoren durch das Proteasom hergestellt werden und durch TAP transportiert werden, müssen sie auf ihre endgültige Größe durch Aminopeptidasen des ER getrimmt werden, bevor sie als Antigene präsentiert werden. Falls dieses Trimmen mit unterschiedlichen Raten für verschiedene Aminosäuren erfolgt, so beein-

Auf diesen Überlegungen beruhend, wurde die Häufigkeit der Aminosäuren A, F, I, L, M, V, W und Y untersucht. Alle diese Aminosäuren begünstigen den TAP Transport und das proteasomale Schneiden, wenn man über die in den Vorhersagen verwendete Region mittelt. Wenn TAP und Proteasom hauptsächlich für die Häufigkeit der Aminosäuren im N-terminalen Processing Motiv verantwortlich sind, ist zu erwarten, dass diese Aminosäuren an allen Positionen in diesem Bereich gehäuft auftreten. Die Werte in Tabelle 15 zeigen deutlich, dass dies nicht der Fall ist. So sind beispielsweise die Aminosäuren W, I und V an der -1 Position geringer häufig anzutreffen, als im Mittel über die Proteine. Sie treten jedoch im größeren Abstand zum N-Terminus (Positionen -4 bis -7) in höherer Frequenz auf. Dieses weist darauf hin, dass das N-terminal Processing Motiv und somit die Selektion der MHC Liganden ebenfalls von der Spezifität der Aminopeptidasen beeinflusst wird und nicht nur von TAP und Proteasom.

		Entfernung zum Epitop N-Terminus																
		-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	N-term	
Korrelation der Abweichungen	Entfernung zum N-Terminus	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1		
		X																
		-0.03	X															
		-0.22	-0.07	X														
		-0.01	-0.06	0.43	X													
		0.46	0.22	-0.18	0.13	X												
		-0.19	0.34	-0.13	0.04	0.48	X											
		0.26	0.05	0.05	0.22	0.01	-0.12	X										
		0.04	0.00	0.31	-0.05	-0.08	0.03	0.10	X									
		0.32	-0.37	0.04	-0.10	-0.26	-0.28	0.40	0.43	X								
		-0.11	-0.34	0.32	-0.01	-0.39	-0.01	0.23	0.48	0.73	X							
		0.22	-0.07	0.25	-0.26	-0.08	-0.22	0.05	0.58	0.64	0.58	X						
		0.17	-0.24	0.44	-0.14	-0.22	-0.34	0.01	0.75	0.46	0.50	0.53	X					
		0.17	0.03	0.04	-0.07	-0.17	-0.32	-0.13	0.32	0.16	0.10	0.26	0.56	X				
		0.12	0.49	0.02	0.07	0.10	0.14	-0.12	-0.01	-0.31	-0.20	0.12	-0.03	0.41	X			
		-0.1	0.31	0.44	-0.14	-0.19	-0.07	-0.28	0.12	0.03	0.02	-0.09	0.30	0.12	0.51	0.76		
	N-term	-0.25	0.60	0.45	0.41	0.03	0.07	-0.13	-0.02	-0.44	-0.28	-0.22	0.01	0.17	0.43	0.25	X	

Unterschiedliche Spezifitäten der cytosolischen und endoplasmatischen Amino-peptidasen beeinflussen das N-terminale Processing Motiv

81

teten Aminosäurehäufigkeiten am N-Terminus von MHC Liganden (Tabelle 15) verglichen und die effektiven Hydrolyseraten an dieser Position berechnet (Abbildung 30). Für um bis zu drei Aminosäuren verlängerte Precursoren von MHC I Liganden findet man höhere Umsatzraten im ER als im Cytosol, noch längere Precursoren werden stärker im Cytosol hydrolysiert als im ER. Eine hohe ER Umsatzrate bewirkt eine effiziente Freisetzung des Epitops, während eine niedrige Umsatzrate im Cytosol es den Precursoren besser ermöglicht von TAP in das ER transportiert zu werden. Zusammengefasst, deuten diese Ergebnisse darauf hin, dass viele der gefundenen MHC I Liganden als längere Precursoren vom Proteasom erstellt wurden, dann von cytosolischen Aminopeptidasen verkürzt wurden, bis nur noch wenige Aminosäuren zusätzlich am N-Terminus sind und anschließend von TAP in das ER transportiert wo die endoplasmatischen Aminopeptidasen das endgültige Epitop generieren. Die Aminosäuren am N-Terminus der Liganden sind im Mittel schlechte Substrate für Aminopeptidasen, was eine relativ hohe Stabilität bewirkt, wodurch die Wahrscheinlichkeit einer Bindung an MHC I Moleküle und damit der Präsentation steigt.

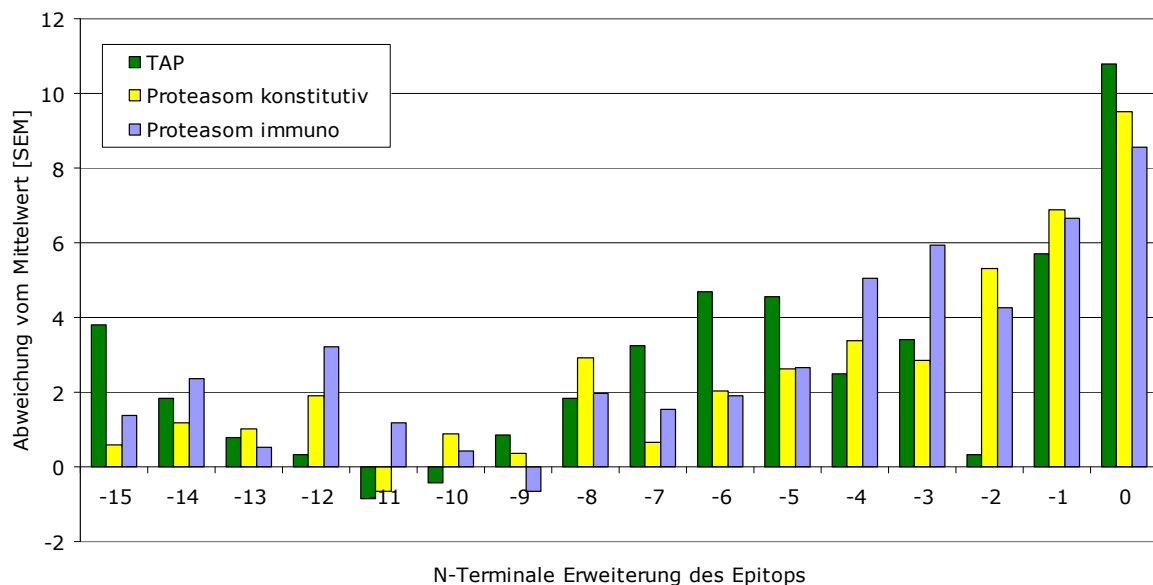


Abbildung 29: Proteasomale Generierung und TAP Transport von MHC Ligand Precursoren

Durchschnittliche TAP Transporteffizienz und proteasomale Schnittraten für N-Terminal verlängerte Precursoren von Epitopen. Die Berechnung der Werte beruht auf einem Sequenzfenster um die Schnittstelle beziehungsweise den N-Terminus. Die Werte sind die Differenz zur entsprechenden mittleren Rate des Prozesses über alle Proteine in Einheiten der Standardabweichung vom Mittelwert.

3.4.3 Diskussion

In diesem Abschnitt wurde der Einfluss des N-terminalen Trimmens auf die Selektion von MHC Liganden untersucht. Hierzu wurden die N-terminalen Verlängerungen von bekannten Liganden analysiert und versucht die beobachteten Häufigkeiten mit Modellen für proteasomale Schnittraten und TAP Präferenzen zu erklären. Die verbleibende Diskrepanz zwischen der beobachteten Verteilung der Aminosäuren und der von TAP und Proteasom Spezifität zu erwartende Verteilung ist auf den Einfluss der Aminopeptidasen im ER und Cytosol zurückzuführen.

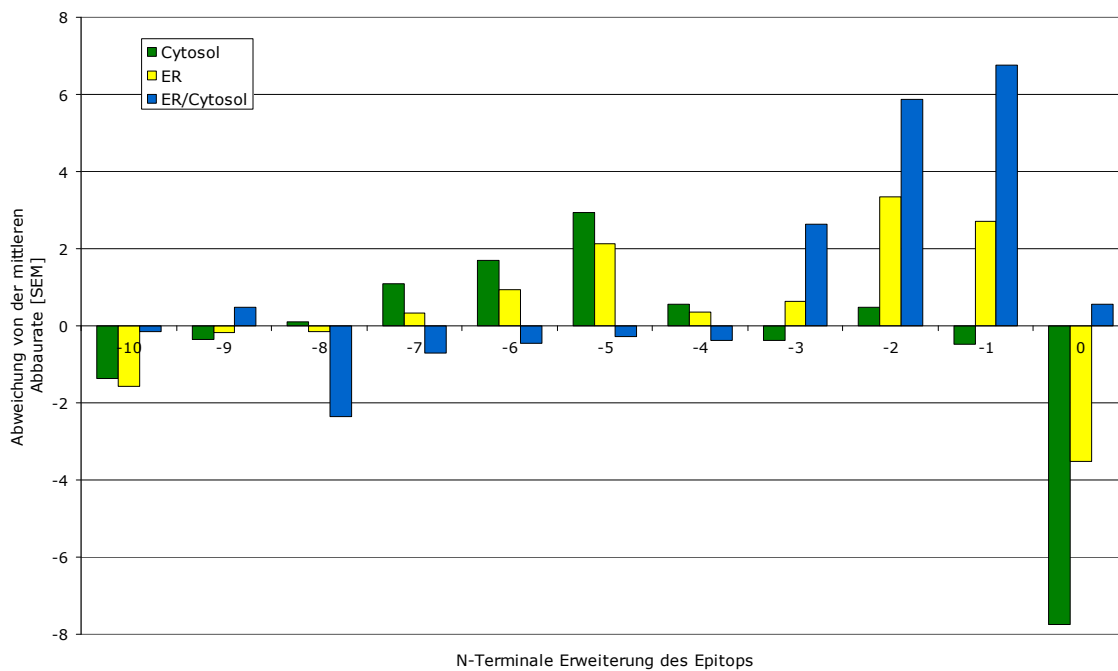


Abbildung 30: Trimmungsraten von MHC Liganden im Cytosol und ER

Durchschnittliche Hydrolyserate von N-terminal verlängerten MHC Liganden. Um bis zu 3 Aminosäuren verlängerte Peptide werden bevorzugt im ER getrimmt, längere Peptide eher im Cytosol. Die tatsächlichen Liganden weisen deutlich verringerte Schnittraten der Aminopeptidasen auf.

Es wurde ein „N-terminales Processing Motiv“ mit einer Länge von etwa acht Aminosäuren identifiziert. Dieses stimmt mit dem Längenbereich der beim Proteinabbau entstehenden Fragmente gut überein. So wird das SIINFEKL Epitop aus Ovalbumin hauptsächlich als, um ein bis sieben Aminosäuren verlängerte, Precursoren im Cytosol gefunden[97]. Weiterhin bevorzugt Tripeptidyl Peptidase II, die zusätzlich zum Proteasom für die Generierung von Precursoren und Epitopen verantwortlich ist[140, 152], Oligo-

peptide mit mehr als 15 Aminosäuren und erzeugt Peptide mit weniger als 16 Aminosäuren[151]. Dadurch werden auch durch TPPII Precursoren mit einem um bis zu acht Residuen verlängerten N-Terminus generiert, da die MHC I Epitope zwischen acht und 11 Aminosäuren lang sind.

Das Proteasom hat eine deutlich höhere Schnittrate in der Region bis zu fünf Aminosäuren distal des Epitop N-Terminus und TAP transportiert Precursoren mit bis zu acht zusätzlichen Aminosäuren besser als zufällige Peptide aus den gleichen Proteinen, aus denen die MHC Liganden entstammen (Abbildung 29). Diese TAP Präferenz wurde mit einem Sequenzfenster von nur drei Aminosäuren gefunden und das Ergebnis stimmt gut mit der berichteten Längenpräferenz von TAP[125] mit hoher Effizienz für 8 bis 16mer überein.

Bestimmung der Aminopeptidaseaktivität und Analyse der beobachteten Aminosäurehäufigkeiten N-terminal der MHC Liganden führt zu der Schlussfolgerung, dass die endoplasmatischen Aminopeptidasen die drei Aminosäuren direkt am Epitop deutlich effizienter abspalten, als die cytosolischen Aminopeptidasen. Dieses suggeriert, dass das finale Trimmen der MHC Liganden vornehmlich im ER stattfindet. Umgekehrt wurde festgestellt, dass die cytosolischen Aminopeptidasen effizienter beim Abtrennen der Residuen an Positionen -8 bis -4 sind, jedoch ist diese Präferenz nicht so stark ausgeprägt. Die Aminosäureverteilung in dieser Entfernung vom Epitop wird also sowohl von cytosolischen als auch endoplasmatischen Aminopeptidasen beeinflusst aber ebenso von den generierenden Prozessen durch das Proteasom oder TPPII.

Das Proteasom hat eine hohe Schnittrate im N-terminalen Processing Motiv und auch TPPII hat aufgrund seiner Längenpräferenz eine hohe Wahrscheinlichkeit in diesem Bereich zu schneiden und daher haben die Aminopeptidasen zunehmend weniger Einfluss auf die Aminosäureverteilung in diesem Processing Motiv, je weiter man vom eigentlichen MHC Liganden entfernt ist.

Diese Ergebnisse führen zu dem in Abbildung 31 dargestellten Modell der Antigenprozessierung. Der erste Schritt ist die Entfaltung des abzubauenen Proteins und die anschließende Generierung von Oligopeptids mit bis zu 25 Aminosäuren Länge, die zum größten Teil schon den korrekten C-Terminus der entstehenden MHC Liganden besitzen. Die durch das Proteasom durchgeführten Schnitte liegen sowohl innerhalb als auch außerhalb des N-terminalen Processing Motivs, wobei jedoch im Mittel die Schnittrate in diesem Bereich höher ist. Peptide länger als 15 Aminosäuren, haben eine hohe Wahrscheinlichkeit von TPPII weiter geschnitten zu werden[32, 151]. So entstehen, zum Teil unter Mithilfe von TPPII, MHC I Epitop Vorläufer mit bis zu acht zusätzlichen Aminosäuren am N-Terminus und dem endgültigen C-Terminus. Precursoren dieser Länge können effizient von TAP in

das ER transportiert werden. Das Trimmen der Aminosäuren an den Positionen -8 bis -4 kann sowohl im Cytosol als auch im ER stattfinden. TPPII kann auch an diesem Prozess beteiligt sein, jedoch wird der Einfluss aufgrund der Längenpräferenz dieses Enzyms immer geringer. Die Aminosäuren an Positionen -2 und -1 werden vornehmlich im ER vom endgültigen MHC Liganden abgespalten. Die geringe Schnittrate der Aminopeptidasen für den tatsächlichen N-Terminus der Epitope ist ein Hinweis darauf, dass das fertige Epitop an das MHC I Molekül bindet, jedoch ist diese Reihenfolge nicht zwingend. Die Bindung an das MHC Molekül schützt das Epitop davor, weiter abgebaut und damit zerstört zu werden. Der entstandene MHC I Komplex wird dann aus dem ER über den vesikulären Transportweg zur Zelloberfläche befördert und präsentiert.

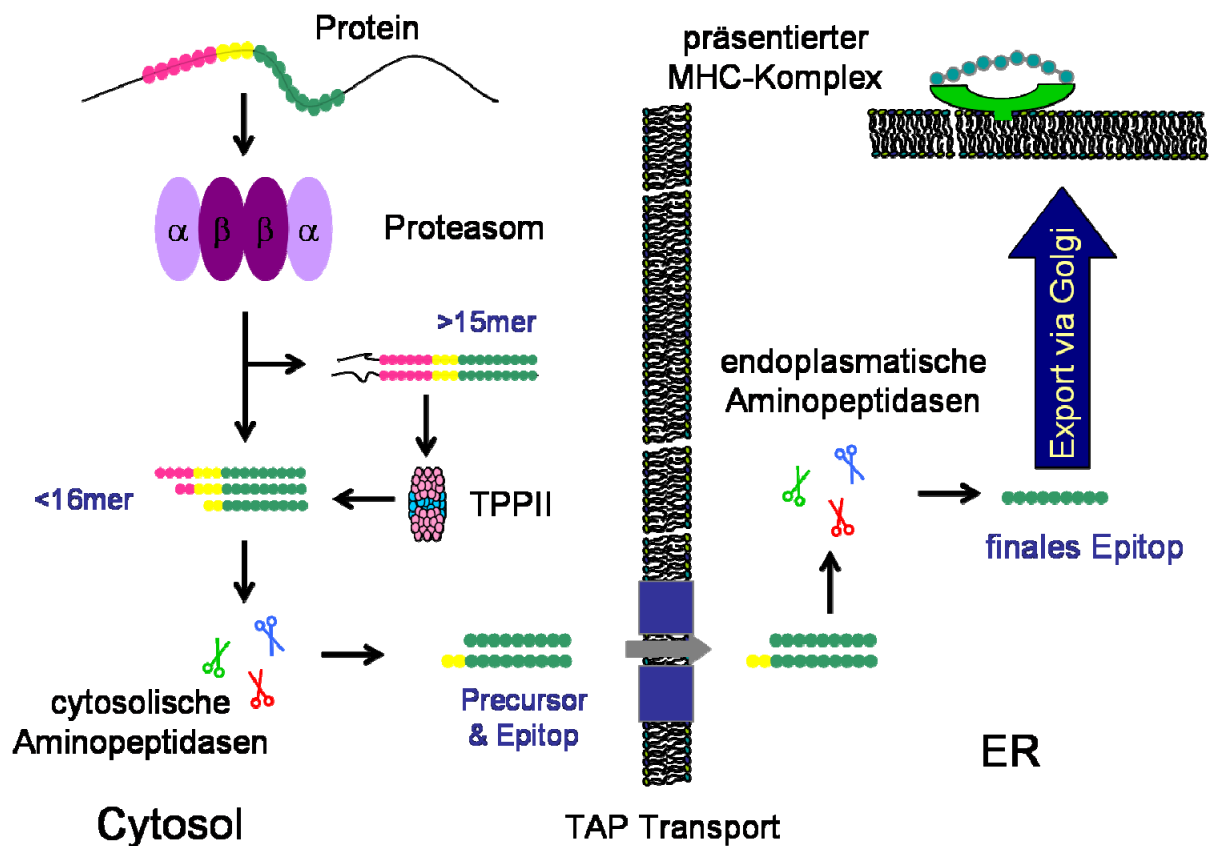


Abbildung 31: Modell der Antigenprozessierung

Die Abbildung wird ausführlich in der Diskussion dieses Abschnitts erläutert.

Abschließend soll hier noch angemerkt werden, dass in diesem Abschnitt häufig auf beobachtete Aminosäurehäufigkeiten Bezug genommen wird. Diese Verteilungen sind letztendlich *in vivo* Daten und stellen somit das Produkt des gesamten MHC I Antigenpräsentationsweges dar. Wenn mit einem Modell, das auf *in vitro* Daten beruht, Teile der beobachteten Verteilungen erklärt werden können, ist das ein starker Hinweis darauf, dass

der entsprechende Prozess für die Generierung der Aminosäurehäufigkeiten verantwortlich ist und somit die Selektivität des Prozesses daraus abgeschätzt werden kann.

Die hier dargestellten Ergebnisse konnten im Journal of Immunology veröffentlicht werden[154].

3.5 Kombination verschiedener Schritte des MHC I Pathways

In diesem Kapitel werden die zuvor beschriebenen Modelle zu einzelnen Schritten des MHC I Pathways kombiniert. Hierbei wird nicht nur auf die Steigerung der Vorhersagequalität Wert gelegt, sondern auch auf die Frage, wie man die Modelle kombinieren sollte und ob die einzelnen Schritte aufeinander abgestimmt sind oder unabhängig voneinander sind.

3.5.1 Sequentielles Verwenden von einem Modell als Filter für das nächste Modell ist bei geeignetem Cutoff möglich.

Ein Ansatz zur Kombination von Vorhersagemodellen zu einzelnen Schritten des MHC Class I Antigen Präsentationsweges ist die Annahme, dass ein Peptid jeden einzelnen Schritt durchlaufen muss und bei jedem Schritt individuell entschieden wird, wie das Peptid weiter prozessiert wird. Somit wirkt jeder Schritt der Antigenpräsentation als Filter auf die Menge der Peptide und nur Peptide, die alle Filter durchlaufen werden präsentiert.

Vorteile dieser Art der Kombination ist die Einfachheit der Kombination und der Entscheidung, ob ein Protein kombiniert wird. Nachteilig ist jedoch, dass man einen Cutoff für jeden Prozess festlegen muss und quantitative Informationen zu den Peptiden verloren gehen.

Bei der Bewertung der Güte der Kombination von verschiedenen Modellen kann eines der Modelle weiterhin eine quantitative Vorhersage liefern, also nicht als binärer Filter fungieren. Die Performance des Gesamtmodells wird dann anhand der ROC Kurve des letzten Modells der Sequenz bestimmt, wobei die Sensitivitäten der Filterprozesse die maximale Sensitivität des finalen Schritts bestimmen.

3.5.2 Die Kombination von zwei Schritten des Pathways mit der Filter Methode verbessert die Vorhersagequalität.

TAP und MHC Vorhersage kombiniert

Wahl eines geeigneten Cutoff für das TAP Modell

Es werden die Parameter für das TAP Modell festgelegt. α ist gleich 0.2 und L gleich 10, da diese Werte auf dem HLA-X Datensatz die besten Ergebnisse erzielten. Ausgehend von Abbildung 24 wird ein Wert als Schranke für den TAP Transport festgelegt, für den entschieden wird, ob ein Peptid von TAP transportiert wird oder nicht. Dieser Wert, für den auf dem HLA-X Datensatz 32 Prozent der Zufallspeptide und nur 1.5 Prozent der Epitope als nicht transportabel eingestuft werden, ist gleich 1.0.

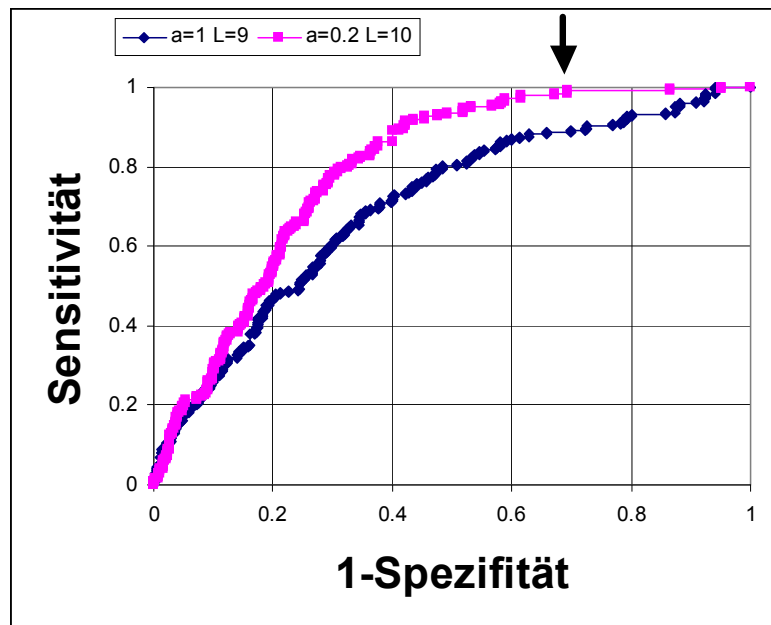


Abbildung 32: ROC Kurve für HLA-X Datensatz

Die Achseneinteilung ist wie in Abbildung 21. In blau ist das ursprüngliche Modell der 9mer Matrix mit einem α von 1 dargestellt. Der AUC = 0.702. In violett ist das erweiterte Modell mit den optimalen Parametern für diesen Datensatz (α gleich 0.2 und L gleich 10) aufgetragen. Der AUC ist hier 0.791. Der Pfeil deutet auf den Cutoff-Wert 1.0, bei dem etwa 98,5 Prozent Sensitivität und 32 Prozent Spezifität erreicht werden.

Ergebnis der Kombination von TAP- und MHC-Modell

Die Vorhersage von Epitopen wird signifikant verbessert, der AUC Wert steigt von 0.919 auf 0.932. In dieser Region ist es sehr schwer, überhaupt Verbesserungen zu erreichen. Hauptsächlich wird die Vorhersage in der Region der ROC Kurve mit hoher Sensitivität verbessert. So kommt man, wenn man eine Sensitivität von 100 Prozent verlangt von einer Spezifität von 52 Prozent zu einer von 62 Prozent. Hierbei ist jedoch zu beachten, dass durch das Filtern mittels TAP 1.5 % der Epitope als nicht transportierbar eingestuft wurden und somit die maximale Sensitivität 98.5 % bleibt.

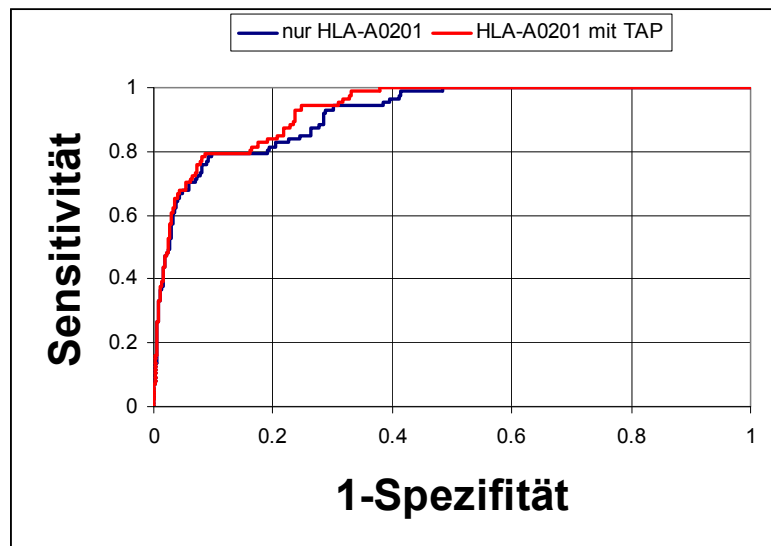


Abbildung 33: ROC Kurve für die Kombination von MHC und TAP Bindungsvorhersagen auf HLA-A0201 Datensatz

Die Auftragung ist wie in Abbildung 21. In blau ist die bereits sehr gute Trennung von Epitopen und Zufallsneunmeren durch die MHC Affinitätsvorhersage[111] auf dem HLA-A0201 Datensatz aufgetragen. Der AUC Wert beträgt 0.919. In rot ist die Kombination der TAP Vorhersage mit der MHC Affinitätsvorhersage dargestellt. Für die TAP Vorhersage werden die auf dem HLA-X Datensatz als optimal bestimmten Parameter (α gleich 0.2, L gleich 10 und Cutoff gleich 1.0) verwendet. Der AUC Wert beträgt hier 0.932. Die Verbesserung der Vorhersagequalität liegt größtenteils im Bereich hoher Sensitivität. Dieses ist Voraussetzung, wenn man mehrere Modelle von verschiedenen Schritten des Pathways kombinieren möchte, da man bei jedem Schritt möglichst keine echten Epitope aus seinem Datensatz entfernen sollte.

Diese Zwei-Schritt Vorhersage wird ebenfalls für alle verschiedenen Maus MHC Allele durchgeführt, für die MHC Matrizen verfügbar sind. Die Epitopvorhersage kann in drei von vier Fällen verbessert werden, aber leider sind hier pro Allel nur sehr wenige Epitope in der Datenbasis enthalten, so dass man bei der Einstufung dieser Verbesserung vorsichtig sein muss. Auch kommt hier zum Tragen, dass die TAP Vorhersage auf dem Maus Datensatz eine geringere Qualität hat als auf dem HLA-X Datensatz, was zum großen Teil in Unterschieden der Substratspezifität des Transporters zwischen den Spezies liegt.

Insgesamt wurde gezeigt, dass die Vorhersage von Epitopen noch verbessert wird, wenn man die Modelle kombiniert. Hier wird TAP als ein Filter implementiert, der Peptide mit sehr geringen Transportwerten nicht mehr in das ER transportiert, so dass sie dann für die MHC Moleküle nicht mehr zur Verfügung stehen.

Proteasom und TAP Vorhersage kombiniert

Die Cutoff Werte der Modelle wurden so gewählt, dass die Sensitivität jeweils 99 % beträgt. Bei der Kombination der Proteasom Modelle verbesserte sich die Spezifität des TAP Modells von 33.6 % auf 36.0 % für das konstitutive Proteasom beziehungsweise 37.4 % für das Immunoproteasom (Tabelle 17). Diese Verbesserung der Modellperformance zeigt, dass es prinzipiell akzeptabel ist die Modelle als Filter zu verwenden.

Tabelle 17: Vorhersagequalität der Kombination von TAP und Proteasom

Modell	Spezifität	
	einzel	kombiniert mit TAP
ProtSMM-c	3.0%	36.0%
ProtSMM-i	6.1%	37.4%
TAP	33.6%	---

Der Cutoff aller Modelle wurde so festgelegt, dass die Sensitivität 99% beträgt. Die Kombination der Proteasommodelle mit TAP verbessert die Gesamtvorhersage.

Kombination von mehr als zwei Teilmodellen des MHC Pathways

Kombination von mehr als zwei Modellen mit der Filter Methode verschlechtert die Güte des Gesamtmodells beziehungsweise es wurde keine Kombination von Cutoff Werten für Proteasom, TAP und MHC gefunden, bei der das Gesamtmodell besser wurde als die MHC Vorhersage alleine. Dieses liegt sicher mit an der hohen Qualität und Selektivität der MHC Modelle, wodurch eine Verbesserung nur in Bereichen sehr hoher Sensitivität erreicht werden kann, da das Herausfiltern von nur wenigen echten Epitopen durch die vorgeschalteten Prozesse sofort zu einer geringeren Sensitivität des Gesamtmodells führt.

Bei Verwendung der Filtermethode verliert man den Vorteil der hier erstellten Modelle, dass für jeden Teilschritt Wahrscheinlichkeiten oder Raten angegeben werden können, da die Aussage der einzelnen Teilmodelle auf eine ja/nein Entscheidung reduziert wird.

3.5.3 Durch additive Kombination der Modelle erhält man relative Peptidmengen.

Die in dieser Arbeit gezeigten Modelle geben quantitative Aussagen, die sich als relative Raten oder Wahrscheinlichkeiten interpretieren lassen, mit der ein Peptid den entsprechenden Prozess durchführt. Bei der Kombination der Modelle als Filter wird diese Eigenschaft auf eine binäre Entscheidung reduziert. An dieser Stelle wird motiviert, dass die Addition der Werte der einzelnen Prozesse ein gutes und immer noch quantitatives Maß

für die Effizienz eines potentiellen MHC Liganden darstellt. Es sind jedoch auch für diese Art der Kombination eine Reihe von vereinfachenden Annahmen zu treffen.

Die Wahrscheinlichkeit, dass ein Peptid an das MHC Molekül bindet, ist proportional zur Konzentration des Peptids im ER und auch zu seiner Affinität.

(18)

$$P_{MHC} = k_1 * [Peptid]_{ER} * \frac{1}{IC_{50,MHC}}$$

Zu beachten ist, dass hier der IC_{50} Wert nicht logarithmiert ist. Die Konzentration des Peptids im ER ist gleich der durch TAP importierten Menge, wenn man Export oder Abbau vernachlässigt beziehungsweise annimmt, dass Abbau und Export für alle Peptide identisch sind und man nur die effektiv zur Verfügung stehende Konzentration betrachtet, die dann proportional zur importierten Peptidmenge ist.

Die von TAP transportierte Menge an Peptid ist proportional zur Affinität des Peptids zu TAP und zur Menge des im Cytosol vorhandenen Peptids.

(19)

$$[Peptid]_{ER} = k_2 * [Peptid]_{Cytosol} * \frac{1}{IC_{50,TAP}}$$

Auch hier ist der IC_{50} Wert nicht logarithmiert.

Die Menge aller Peptide mit einem bestimmten C-Terminus im Cytosol ist proportional zur proteasomalen Schnittrate für diesen C-Terminus.

(20)

$$\left\{ [Peptid]_{Cytosol} \right\}_{CTerm} = k_3 * P_{Proteasom}(CTerm)$$

Die Menge eines einzelnen Peptids dieser Menge ist zusätzlich noch von der Länge des Peptids und seiner Sequenz abhängig.

Die Menge aller Peptide im ER die einen MHC Liganden oder einen N-terminal verlängerten Vorläufer bilden lassen sich durch Kombination von Gleichungen 19 und 20 ausdrücken. Kürzere Peptide als das Epitop, die potentiell vom Proteasom erzeugt werden und auch längere Peptide, als von TAP effizient transportiert werden, werden hierbei vernachlässigt beziehungsweise angenommen, dass der Anteil dieser Peptide an den Peptiden mit einem bestimmten C-Terminus konstant ist.

(21)

$$\{[Peptid]_{ER}\}_{CTerm} = k_4 * P_{Proteasom}(CTerm) * \frac{1}{\langle IC_{50,TAP} \rangle}$$

Hierbei ist $\langle IC_{50,TAP} \rangle$ der effektive IC_{50} Wert für diese Menge von Peptiden, dessen Logarithmus mittels Gleichung 15 berechnet wird. Die Motivation erfolgte in Abschnitt 3.2.5 Gleichungen 10 bis 15. In das TAP Modell (betrachtete Länge $L=10$, $\alpha=0.2$) sind hierbei schon Annahmen über die Peptide eingeflossen. Es zeigt sich, dass das starke Gewichten des C-Terminus durch α auf sich gegenseitig ergänzende Präferenzen des Proteasoms und TAP an der P1 Position zurückzuführen ist. In der Kombination von TAP- und Proteasomscores erhält man bessere Ergebnisse in der MHC Liganden Vorhersage, wenn $\alpha=1$ gewählt wird. Dieses ist leicht verständlich für das Immunoproteasom, bei dem die Korrelation der Aminosäurepräferenz an der P1 Position mit der C-terminalen Präferenz von TAP relativ hoch ist ($r^2=0.37$). Für das konstitutive Proteasom ist diese Korrelation praktisch nicht existent ($r^2=0.02$) aber trotzdem ergänzen sich TAP und Proteasom in der Selektion der Peptide derart, dass ein Wichten der C-terminalen Vorhersage von TAP in der Kombination nicht mehr zu verbesserten Ergebnissen führt. Die maximal betrachtete Länge von Peptiden reflektiert ebenfalls die Längenverteilung der vom Proteasom produzierten Peptide. So scheint die Fraktion der noch weiter verlängerten Peptide nicht groß genug zu sein, um die Vorhersage von MHC Liganden deutlich zu verbessern. Dieses heißt jedoch nicht, dass diese Peptide in keinem Fall transportiert werden.

Diese Ausführungen zeigen, dass es nicht einfach ist die einzelnen *in vitro* Modelle zu kombinieren ohne sie auf Eigenschaften des gesamten Pathways abzustimmen. Hier wurde motiviert, dass die Addition der logarithmierten Werte der einzelnen Vorhersagemethoden einen geeigneten Schätzer für die Effizienz eines potentiellen MHC Liganden liefert. Die Verwendung eines mittleren Wertes für die TAP Transporteffizienz ist hierbei eine Kompensation für die Unsicherheit, welche Fragmente vom Proteasom produziert werden. Die Bestimmung des Parameters L wurde anhand von beobachteten MHC Liganden durchgeführt, also an *in vivo* Daten trainiert.

Kombination der Vorhersage von TAP und Proteasom

Die Bewertung eines potentiellen MHC Liganden erfolgt also, indem der Ligand und die ihn umgebenden Aminosäuren mit den Modellen für TAP und Proteasom bewertet werden. Diese Scores werden dann additiv kombiniert (Gleichung 22). Diese Scores sind im

Wesentlichen proportional zum Logarithmus der Menge des Liganden und seiner N-terminal verlängerten Vorläufer im ER.

(22)

$$Score_{Ligand} = \log(P_{\text{Proteasom}}(Ligand)) + \log(\langle IC_{50,TAP}(Ligand) \rangle)$$

Die Vorhersagequalität verbessert sich bei der Kombination der Modelle für TAP und das Proteasom deutlich (Abbildung 34 und Tabelle 18). Bemerkenswert ist, dass trotz der großen Unterschiede in der Vorhersagequalität der beiden Proteasommodelle in der Kombination mit TAP fast die gleichen Ergebnisse erzielt werden. Offensichtlich produziert das Immunoproteasom vermehrt Peptide, aus denen ein MHC Ligand entstehen kann, während das konstitutive Proteasom nicht so spezifisch ist.

Das hier präsentierte Modell für TAP und Proteasom für die Vorhersage von Epitopen hat bei einer Spezifität von 50 % eine Sensitivität von 97 % und ist unabhängig vom in den Zellen exprimierten MHC Allel. Somit kann mit dieser Methode die Hälfte der potentiellen MHC Liganden eliminiert werden, ohne mehr als 3 % der echten Epitope zu verlieren.

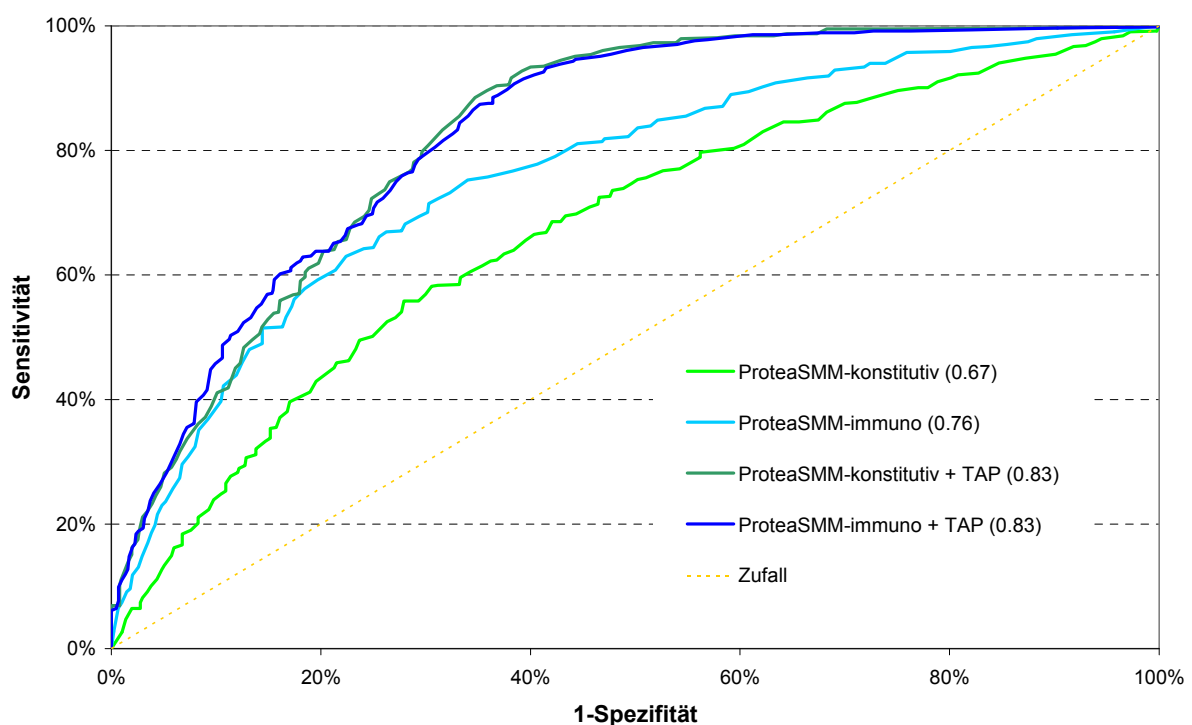


Abbildung 34: ROC von ProteaSMM in Kombination mit TAP

Die verwendeten Methoden für die Vorhersage von MHC Liganden sind in der Legende angegeben. Die MHC-Liganden werden aus den sie enthaltenden Proteinen identifiziert. Der Datensatz ist in Abschnitt 3.3.1 erläutert.

Tabelle 18: AUC Werte für ProteaSMM in Kombination mit TAP

Datensatz	ProteaSMM-konstitutiv	ProteaSMM-immuno	nur TAP	TAP +	
				ProteaSMM-konstitutiv	ProteaSMM-immuno
RCC-c	0.684	0.755	0.818	0.825	0.831
RCC-i	0.668	0.785	0.844	0.857	0.862
Komplett	0.674	0.761	0.814	0.827	0.831

Kombination von Proteasom, TAP und MHC Vorhersage

Die Kombination der Scores für TAP, Proteasom und MHC ergibt sich analog zu Gleichung 22 nach:

(23)

$$Score_{Ligand} = \log(P_{\text{Proteasom}}(Ligand)) + \log(IC_{50,TAP}(Ligand)) + \log(IC_{50,MHC}(Ligand))$$

Der hier erhaltene Wert ist proportional zum Logarithmus der Effizienz des Liganden für den gesamten Pathway.

Es wurden mehrere ARB Matrizen⁹ verwendet für die Vorhersage von humanen MHC Molekül Affinitäten, die von der Sette Gruppe für eine Vielzahl von MHC Allelen publiziert wurden[155, 156, 157, 158, 159]. Alle verwendeten Matrizen basieren auf *in vitro* Daten. Diese Matrizen wurden so normalisiert, dass sie logarithmierte relative IC₅₀ Werte generieren. Es wurden nur Matrizen für Allele verwendet, die mit mindestens 10 Liganden im SYF-human Datensatz (Abschnitt 3.3.1) enthalten sind. Der durchschnittliche AUC Wert für die MHC Vorhersage beträgt 0.965, was ein sehr hoher Wert ist und nochmals unterstreicht, dass die Bindung an das MHC Molekül der am stärksten selektive Schritt bei der Antigenpräsentation ist und auch das die Modelle für diesen Schritt eine sehr hohe Qualität haben. In Kombination mit der Vorhersage von TAP und Proteasom kann die Vorhersagequalität noch gesteigert werden auf AUC=0.971 sowohl für das konstitutive als auch das immuno Proteasom (Tabelle 19).

Vergleich der kombinierten Vorhersage mit anderen Pathway Methoden

Um die Qualität des hier präsentierten MHC I Pathway Modells zu testen, soll es mit SYFPEITHY[106] und BIMAS[160] verglichen werden. Diese Methoden wurden zumindest zum Teil auf den in SYFPEITHI enthaltenen Daten trainiert worden, wodurch der SYF-human nicht für einen objektiven Vergleich verwendet werden kann. Stattdessen kann

⁹ ARB Matrix: **a**verage **r**elative **b**inding, bezeichnet eine Matrix die nach dem Prinzip der unabhängigen Bindung aufgebaut wurde.

aber der RCC Datensatz verwendet werden, da er nur zuvor unbekannte MHC Liganden enthält und nicht für das Training einer der Methoden verwendet worden ist.

Der Vergleich der Methoden zeigt, dass die Kombination der drei *in vitro* Modelle zu mindestens die gleiche prädiktive Qualität hat, wie die beiden anderen etablierten Modelle, die auf *in vivo* Daten trainiert worden sind (Abbildung 35). Die Vorhersage mit der ProteaSMM-i Methode ist in drei der vier getesteten Fälle besser als ProteaSMM-c. Für das eine Allel, bei der die Vorhersage mit ProteaSMM-c besser ist, ist die reine MHC Vorhersage nochmals deutlich besser, was wahrscheinlich darauf zurückzuführen ist, dass viele der vom HLA A-03 präsentierten MHC Liganden P1K Liganden sind und wahrscheinlich nicht vom Proteasom hergestellt werden.

Tabelle 19: AUC Werte für Kombination von TAP, Proteasom und MHC Vorhersage

SYFPEITHI-MHC Liganden	nur MHC	MHC + TAP +	
		ProteaSMM-c	ProteaSMM-i
HLA-A01	0.992	0.997	0.997
HLA-A0201	0.951	0.956	0.955
HLA-A24	0.951	0.958	0.958
HLA-B07	0.989	0.993	0.993
HLA-B5101	0.943	0.952	0.951
RCC Datensatz	nur MHC	MHC + TAP +	
		ProteaSMM-c	ProteaSMM-i
HLA-A01	0.991	0.996	0.996
HLA-A03	0.979	0.977	0.976
HLA-B18	0.995	0.997	0.997
HLA-B40	0.993	0.995	0.995

Ein weiterer Vorteil der hier vorgestellten Methode ist, dass sie auf alle menschlichen MHC Allele anwendbar ist, für die eine Affinitätsmatrix existiert, während für SYFPEITHI und BIMAS eine Mindestanzahl an bekannten Epitopen notwendig ist.

Die Vorhersage von Epitopen aus der Sequenz wurde als eine in Python[161] programmierte Anwendung online zur Verfügung gestellt: mhc-pathway.net[87, 140]. Diese Anwendung wurde mittlerweile weiterentwickelt und in die Internetpräsenz des La Jolla Institute for Allergy & Immunology integriert[114, 162].

3.5.4 Kombination der Modelle mit N-terminaler Trimmung von Peptiden führt zu keiner Verbesserung der Vorhersagequalität

Es war nicht möglich die Gesamtvorhersage von MHC Liganden signifikant zu verbessern, wenn das Modell des N-terminalen Trimmens durch Aminopeptidasen in die Bewertung mit eingeschlossen wurde. Dieses zeigt zum einen, dass es nur schwer möglich ist, die extrem gute Vorhersage durch die MHC Modelle noch zu übertreffen und dass die Kombi-

nation von im Prinzip aufeinander abgestimmten statistischen Modellen stark von den getroffenen Annahmen abhängig ist. Konkurrierende Prozesse, wie das cytosolische Trimmen von Peptiden und der Transport dieser Peptide in das ER lassen sich nur in einem statistischen Modell kombinieren, wenn die relative Stärke dieser Prozesse zueinander aus *in vivo* Daten abgeschätzt wird.

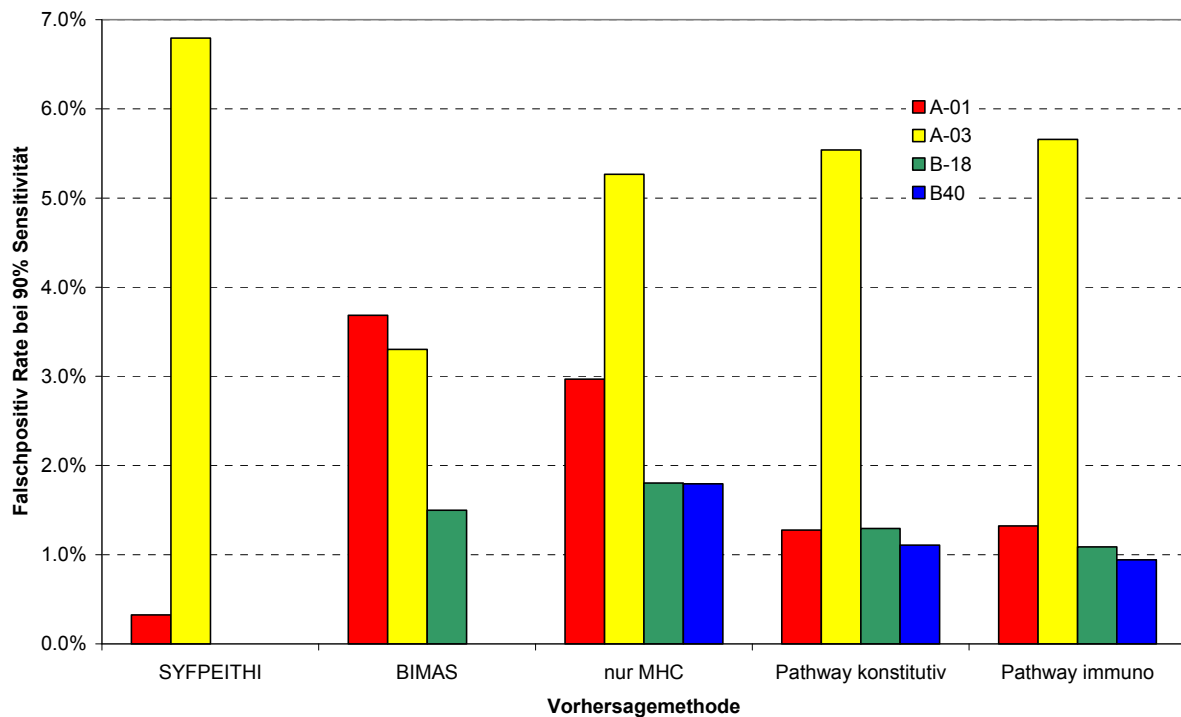


Abbildung 35: Methodenvergleich für Vorhersage von antigenen Peptiden

Es ist die Rate der falsch positiv vorhergesagten Epitope ($1 - \text{Spezifität}$) bei 90% Sensitivität für verschiedene Vorhersagemethoden angegeben. Die einzelnen Balken sind die Ergebnisse für vier verschiedene MHC Allele. Die Methode „nur MHC“ beinhaltet nur die Vorhersage der MHC Affinität, „Pathway konstitutiv“ und „Pathway immuno“ beinhalten die kombinierte Vorhersage von MHC Bindung, TAP Transport und proteasomaler Erstellung des C-terminalen Schnitts, wobei immuno und konstitutiv auf die verwendete proteasomale Methode verweist.

3.5.5 Diskussion

In diesem Abschnitt wurden Möglichkeiten der Kombination der Modelle für die einzelnen Teile des MHC I Präsentationswegs analysiert und durchgeführt. An dieser Stelle sollen einige der erzielten Ergebnisse noch weiter erörtert werden.

Reihenfolge der Kombinationen und verwendete Daten entsprechen der zeitlichen Entwicklung der Modelle.

Die einzelnen hier vorgestellten Modelle wurden innerhalb mehrerer Jahre entwickelt. Dabei änderten sich natürlich die zur Verfügung stehenden Daten wie zum Beispiel die bekannten MHC Liganden in SYFPEITHI. Auch die verfügbaren Modelle von anderen Arbeitsgruppen haben sich in der Zeit verändert.

In dieser Arbeit wurde bei der Darstellung die Reihenfolge gewählt, in der die Modelle entwickelt wurden und bei dem Trainieren und Testen der Modelle wurden jeweils die zum damaligen Zeitpunkt verfügbaren Daten und Modelle ausgewählt und verwendet[87, 140, 154].

Auch kleine Änderungen bei hohen AUC Werten bedeuten starke Änderungen der Vorhersagequalität.

AUC Werte sind ein gutes Maß für die Bewertung von Vorhersagealgorithmen, jedoch sind sie nicht linear und auch Änderungen am AUC Wert sind nicht unabhängig davon, von welchem Wert man startet. So stellt sich die Frage, ob eine Änderung des AUC um 0.005 wie bei der Kombination der MHC Vorhersage für HLA A-01 mit der Vorhersage für Proteasom und TAP (Tabelle 19) eine echte Verbesserung darstellt. Wenn man mit einer Sensitivität von 90 % die Epitope finden möchte, hat die reine MHC Vorhersage eine Spezifität von 97 %, die kombinierte Vorhersage eine Spezifität von 98.8 %. Dieses bewirkt für ein hypothetisches Protein mit 1000 Aminosäuren, dass man statt zuvor 30 Kandidaten für MHC Liganden nur noch 12 experimentell testen muss um die MHC Liganden aus diesem Protein mit 90 % Sensitivität zu finden. Der experimentelle Aufwand wird um 60 % reduziert, jedoch steigt der AUC Wert nur von 0.992 auf 0.997.

Dieses zeigt auch, dass es schwer ist, die schon sehr guten MHC Bindungsvorhersagen zu verbessern, da das fälschliche Ausschließen von einzelnen wahren MHC Liganden durch die zusätzlichen Modellkomponenten fast automatisch zu einer Reduktion der Vorhersagequalität führt.

Das Einschließen des N-terminalen Trimmens von Peptiden verbessert nicht die Vorhersage der MHC Liganden.

Ein gutes Beispiel dafür, dass das Einbeziehen eines Prozesses mit einer nicht ausreichenden Modellqualität die Vorhersagekraft des Gesamtmodells senkt, sind die Modelle für die Vorhersage C-terminaler Schnitte durch das Proteasom. Bei der Entwicklung des Modells für TAP waren diese Modelle noch auf einem schlechteren Stand und in Kombina-

tion verringern sie die Vorhersagequalität der MHC Bindungsvorhersage. Jede dieser Vorhersagemethoden grenzt die Menge an potentiellen Kandidaten für MHC Liganden ein, jedoch sind sie nicht gut genug, um in Kombination mit anderen Komponenten des Pathways zu einer Verbesserung zu führen.

Auch das Einbeziehen der Aminopeptidaseaktivität in die Vorhersage von MHC Epitopen führte zu einer Verringerung der Qualität, obwohl klar gezeigt werden konnte, dass die beobachteten Verteilungen von Aminosäuren N-terminal der bekannten MHC Liganden sich zum Teil mit der Aktivität der Peptidasen erklären lässt. Ursache hierfür ist zum einen, dass die Messung der Aktivität mit sehr kurzen fluorogenen Substraten durchgeführt wurde, die möglicherweise nicht ausreichend repräsentativ für natürliche Peptide sind[154]. Es wurde von experimentellen Partnern hierbei an einem Beispiel gezeigt, dass längere Peptide andere Abbauraten haben aber abhängig von der N-terminalen Aminosäure die gleiche Rangfolge der Aktivitäten zu beobachten ist, wie bei den kurzen Substraten. In der Arbeit von Hearn et al. [163], die das Trimmen im ER genauer untersucht haben, wurden die Ergebnisse der Arbeit von Schatz et al. [154] qualitativ bestätigt. Abweichungen wurden vor allem für geladene N-terminale Aminosäuren gefunden und für Prolin wurde ebenfalls eine Hydrolyseaktivität festgestellt, auch wenn sie geringer ist als bei allen anderen Aminosäuren. Somit kann man festhalten, dass die Wahl des Substratgerüsts die Aktivität beeinflusst und somit die Qualität der Modelle davon abhängig ist. In der Arbeit von Kawahara et al.[150] wurde festgestellt, dass TPPII lange Precursoren trimmt. Dieser Einfluss konnte in der Bestimmung der Aktivitäten durch fluorogene Substrate nicht festgestellt werden, da nicht genügend N-terminale Aminosäuren für TPPII vorhanden waren

Weiterhin stellt sich beim Einbeziehen der Peptidaseaktivität in das Gesamtmodell wiederum die Frage der Art der Kombination. Die Peptidasen arbeiten in Konkurrenz zu den Prozessen des Transports in das ER und der Bindung an MHC Moleküle, wodurch zumindest die relative Stärke der Kombination bekannt sein müsste, um sie mit einzubeziehen. Somit verlangt die Kombination dieser Prozesse im Prinzip ein mechanistisches Modell des Vorgangs, wie es im Kapitel 5 für ein Beispiel durchgeführt wird.

Die Entwicklung und Kombination von in vitro Modellen bringt mehr Erkenntnisse als MHC Liganden Vorhersage mit in vivo Modellen.

Es konnte hier gezeigt werden, dass die kombinierten *in vitro* Modelle mindestens die gleiche prädiktive Qualität haben, wie auf *in vivo* Daten basierende Modelle. Somit stellen die kombinierten *in vitro* Modelle keinen Nachteil bei der Vorhersage von MHC Liganden dar und aus der notwendigen Art der Kombination lassen sich weitere Informationen ge-

winnen. So ist das Verwenden der einzelnen Modelle als Filter nicht so gut, wie die Summe der Scores. Dieses zeigt, dass sich die Raten der einzelnen Prozesse des MHC I Pathways gegenseitig kompensieren können und nicht nur anhand eines Schrittes die Entscheidung fällt, ob ein Peptid präsentiert werden kann. Auch die relative Selektivität der einzelnen Schritte lässt sich nur anhand von *in vitro* Modellen bestimmen.

Weiterhin wurde mit dem *in vitro* Modell des Transports von Peptiden in das ER mittels TAP festgestellt, dass endoplasmatisches Trimmen notwendig ist, um die beobachtete Verbesserung der Vorhersage bekannter MHC Liganden durch Einbeziehen des Transports von Precursoren zu erklären.

Die Häufigkeit von P1K Liganden lässt sich nicht durch proteasomale Generierung erklären, was ein starker Hinweis auf das generelle Herstellen eines Teils der präsentierten Antigene durch andere Endopeptidasen als dem Proteasom ist. Dieses kann durch *in vivo* Modelle nicht erfasst werden.

Zwischen der Vorhersage der kombinierten *in vitro* Modelle und den beobachteten MHC Liganden bestehen keine weiteren signifikanten Abweichungen. Dieses zeigt, dass die bekannten und modellierten Prozesse im Wesentlichen ausreichen, um die Selektivität des Pathways zu erklären und andere alternative Pathways[67, 69] nicht notwendig sind. Dieses ist natürlich kein Beweis für die Nichtexistenz dieser Alternativen, sondern lediglich dafür, dass diese Wege entweder nur geringfügig an den präsentierten Antigenen beteiligt sind oder eine ähnliche Selektivität aufweisen wie das klassische System. Geringfügig bezieht sich hierbei auf die Quantität der beobachteten MHC Liganden. Physiologisch kann eine solche Alternative natürlich den Unterschied zwischen der erfolgreichen Evasion der Immunüberwachung durch ein Pathogen und der wirksamen Bekämpfung sein.

Das Verwenden der einzelnen Schritte als Filter für die Gesamtvorhersage ist sinnvoll, wenn man an einer kurzen Liste von MHC Kandidaten interessiert ist.

Oft ist es nicht notwendig, fast alle Antigene aus einem Protein zu identifizieren, sondern man ist eher an einer möglichst kurzen Liste an Kandidaten interessiert. In diesen Fällen, erhält man auch bei der Kombination der einzelnen Modelle des MHC Pathways als Filter akzeptable Ergebnisse, auch wenn die mittels AUC gemessene Gesamtvorhersagequalität geringer ist. So hat sich gezeigt, dass es ein geeignetes Vorgehen ist, wenn die potentiellen MHC Liganden folgendermaßen selektiert werden: (i) Auswählen der besten Liganden anhand der MHC Bindungsvorhersage. (ii) Alle Kandidaten, die sehr schlecht vom Proteasom hergestellt werden, werden eliminiert. (iii) Alle potentiellen Epitope, die mit geringer Effizienz von TAP transportiert werden und für die kein N-terminal verlängerter Precursor

mit akzeptabler TAP Effizienz existiert, werden ebenfalls eliminiert. Die Reihenfolge ist hierbei nicht festgelegt. Wie restriktiv man bei den einzelnen Schritten vorgeht, hängt davon ab, welcher experimentelle Aufwand für die Validierung der Vorhersagen akzeptiert werden kann. Die meisten Kandidaten sollten bei der Vorhersage der MHC Bindung entfernt werden.

4 Mechanistisches Modell der proteasomalen Fragmentgeneration

Schnittstellenvorhersage versus Fragmentvorhersage

Die Vorhersage von Fragmenten ist nicht identisch mit der Vorhersage von Einzelschnittereignissen, wie sie in Abschnitt 3.3 dargestellt wurde. Die Annahme, man erhält eine akzeptable Fragmentvorhersage, wenn man in der Sequenz eines Proteins immer dann schneidet, wenn man mit der Schnittstellenbewertung über einem vorgegebenen Grenzwert liegt, ist falsch. Zum einen handelt es sich bei der Schnittstellenvorhersage immer um eine Rate beziehungsweise Wahrscheinlichkeit mit der diese Bindung gespalten ist und es ist somit keine deterministische Vorhersage und Spaltung einzelner Schnittstellen in einem Protein ist auch nicht unabhängig von den anderen Schnittstellen.

Die Rate oder Wahrscheinlichkeit mit der ein Fragment erstellt wird, ergibt sich aus dem Produkt der Wahrscheinlichkeiten an beiden Rändern zu schneiden sowie der Wahrscheinlichkeit keinen Schnitt im Inneren des Fragments durchzuführen. Hinzu kommen aber auch noch modifizierende Effekte, die durch den Mechanismus der proteasomalen Fragmenterstellung bedingt sind und von der Schnittstellenvorhersage nicht erfasst werden. Dieses sind unter anderen zusätzlich zu den zuvor diskutierten Einflüssen mehrerer Aminosäuren zusammen (beispielsweise Paarkoeffizienten) auch das Schnittverhalten des Proteasoms an den Rändern des abzubauenen Peptids und die Regulation der Aufnahme von Peptiden in das Proteasom, was das weitere „Verdauen“ von zuvor erzeugten Fragmenten mit einschließt.

Hier soll der Versuch unternommen werden aus der zuvor dargestellten Schnittstellenvorhersage eine Fragmentvorhersage zu entwickeln, die das Verhalten des Proteasoms mit möglichst einfachen Annahmen wiedergibt. Da über den tatsächlich mechanistisch ablaufenden Vorgang der Fragmenterstellung wenig bekannt ist, sind die hier getroffenen Annahmen rein deskriptiv und haben wahrscheinlich keine mechanistische Entsprechung.

4.1 Entwicklung des „slide & slice“ Modells

4.1.1 Das Erzeugen der richtigen Längenverteilung

Die Längenverteilung der Peptide ist Zeitabhängig

Die Längenverteilung der erzeugten Peptide in einem proteasomalen *in vitro* Verdau ist eine Funktion der Zeit. Zum Start des Experiments liegen nur unverdaute Peptide vor, die

mit fortschreitender Zeit zu immer kürzeren Fragmenten abgebaut werden, bis nach entsprechend langer Zeit nur noch die einzelnen Aminosäuren vorhanden sind. Dieser hypothetische Endpunkt wird experimentell nie erreicht, obwohl jede Peptidbindung eine gewisse Wahrscheinlichkeit größer Null hat, vom Proteasom geschnitten zu werden.

Im *in vivo* System verhält es sich hingegen so, dass effiziente Aminopeptidasen alle freigegebenen Peptide abbauen, die nicht lang genug sind, um sich spontan so zu falten, dass sie für die Peptidasen unangreifbar sind. Die Freisetzung von Fragmenten aus dem Proteasom und deren Wiederaufnahme wurden experimentell beobachtet[32] jedoch ist nicht bekannt, in welchem Ausmaß dies geschieht.

Hier wird ein Modell entwickelt, das die Peptidverteilung nach einem langfristigen Verdau (50 Prozent verdautes Substrat, etwa 48 Stunden [130]) reproduziert.

Parametrisieren der experimentellen Daten

Experimentell ist eine log-normale Längenverteilung der erzeugten Peptide beobachtet worden[164]. Es wurde eine entsprechende Verteilungsfunktion an die experimentellen Daten angepasst und die relativen Häufigkeiten der einzelnen Peptidlängen daraus interpoliert (Abbildung 37).

In dem hier entwickelten Modell wird die explizite Aufnahme der Peptide in das Proteasom sowie das weitere Abbauen bereits erzeugter Fragmente vernachlässigt. Ein Modell, das die Aufnahme, Abgabe und Prozessierung von Peptiden als separate Bestandteile beinhaltet existiert [165] und wird im Diskussionsteil dieses Abschnitts weiter erörtert.

4.1.2 Beschreibung des „slide & slice“ Modells

Modellannahmen und Mechanistische Beschreibung der Prozesse

Die Aufnahme des Substrats in das Proteasom wird nicht beschrieben. Es wird davon ausgegangen, dass das Substrat in gestreckter Konformation im Proteasom ist beziehungsweise jede Bindung im Prinzip für die Hydrolyse zugänglich ist. Die insgesamt sechs katalytisch aktiven Zentren des Proteasoms werden zu einem einzigen vereinfacht.

Das Modell beinhaltet zwei mögliche Ereignisse. Zum einen kann sich das Substrat um eine Aminosäure weiter in das Proteasom bewegen oder es kann am aktiven Zentrum geschnitten werden. Für die Bewegung gibt es nur die vorwärts Richtung und die Geschwindigkeit der Bewegung ist Abhängig von der Anzahl der Aminosäuren, die schon jenseits des aktiven Zentrums sind wobei es immer eine Mindestrate der Vorwärtsbewegung hat (Gleichung 24).

(24)

$$v = v_1 + \frac{v_2}{1 + \left(\frac{L}{L_c}\right)^n}$$

Die Variable v ist hierbei die Rate der Vorwärtsbewegung, L entspricht der Anzahl der Aminosäuren des Substrates, die über das aktive Zentrum hinaus sind. Die Parameter v_1 und v_2 sind Geschwindigkeitsraten, wobei v_1 die Basalgeschwindigkeit ist und v_2 die zusätzliche Geschwindigkeit ohne Aminosäuren jenseits des katalytischen Zentrums ist. L_c stellt die kritische Länge und n die Steilheit der Kurve dar, mit der die Geschwindigkeit v von v_1+v_2 auf v_1 absinkt.

Für die Schnittrate wird das in Abschnitt 3.3 entwickelte Modell verwendet. Nach einem Schnittereignis wird das gebildete Fragment direkt freigegeben und das verkürzte Substrat bildet das neue Substrat wobei L wieder auf Null gesetzt wird.

Das Modell ist schematisch in Abbildung 36 dargestellt.

Algorithmische Implementierung

Die Simulation wird als stochastische Monte Carlo Simulation nach dem Gillespie Algorithmus[166] durchgeführt. Das System wird in einem Ausgangszustand gestartet und durchläuft eine Reihe von Zuständen, bis die Simulation beendet ist. Gestartet wird das System mit dem Substrat direkt vor dem aktiven Zentrum des Proteasoms. Dann kann das Substrat in jedem Schritt der Simulation sich entweder um eine Aminosäure vorwärts bewegen oder geschnitten werden. Welches Ereignis durchgeführt wird, wird zufällig jedoch entsprechend der Raten der beiden Prozesse entschieden. Die Simulation ist beendet, wenn das Substrat komplett über das aktive Zentrum bewegt wurde. Um eine Verteilung von Fragmenten zu erhalten, wird diese Simulation entsprechend häufig wiederholt.

4.1.3 Anpassung der Modellparameter an experimentelle Daten

Das Anpassen von Parametern an experimentelle Daten ist in stochastischen Systemen ungleich schwerer beziehungsweise aufwendiger als in deterministischen Systemen. Da bei zwei Durchführungen der gleichen Simulation mit identischen Parametern unterschiedliche Ergebnisse entstehen, ist es nicht möglich mit Standardmethoden zu arbeiten. Eine Optimierung mit dem Metropolis Algorithmus und „simulated annealing“ oder mit genetischen Algorithmen ist möglich.

Hier sind nur wenige Parameter zu optimieren und so wurde der sehr einfache Ansatz gewählt die Parameter innerhalb eines Intervalls zufällig zu wählen und das System damit zu simulieren. Beim Verwenden von ausreichend vielen Parametertupeln erhält man so eine gute Annäherung an das zu erreichende Optimum und weiterhin erhält man so gleichzeitig eine Sensitivitätsanalyse.

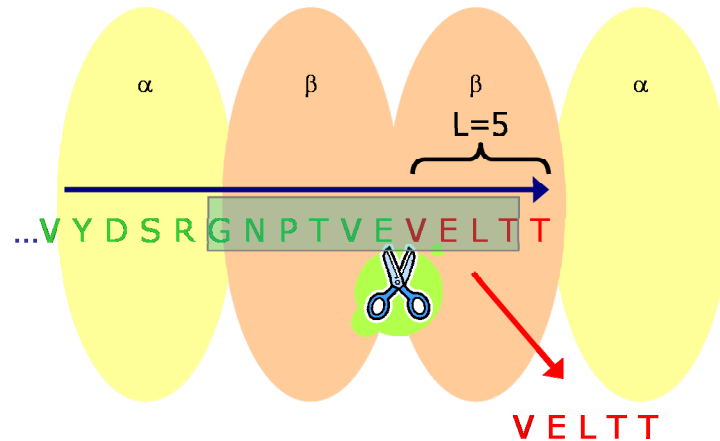


Abbildung 36: Mechanistisches Modell der proteasomalen Fragmentbildung

Das Peptid ist im Proteasom und bereits fünf Aminosäuren sind über das aktive Zentrum hinaus bewegt worden. An dieser Stelle wird die Entscheidung getroffen, ob das Peptid um eine weitere Position nach vorne bewegt wird (blauer Pfeil) oder ob ein Schnitt durchgeführt und das gebildete Fragment freigesetzt wird (roter Pfeil). Die Rate für den Transport ist von der Anzahl der Aminosäuren jenseits des aktiven Zentrums ($L=5$) abhängig. Die Schnittrate wird von den zehn Aminosäuren um das aktive Zentrum (das Fenster von sechs + vier Aminosäuren ist mit einem Kasten hinterlegt) determiniert.

4.2 Ergebnis

4.2.1 Das Modell gibt die experimentell beobachtete Längenverteilung gut wieder

In Abbildung 37 ist die Übereinstimmung der Längenverteilungen von Experiment und Simulation dargestellt. Das hier verwendete sehr einfache Modell ist in der Lage die gewünschte Längenverteilung zu erzeugen.

Die Anpassung der Parameter ist in Abbildung 38 dargestellt. Es ist zu erkennen, dass es für v_1 ein sehr scharfes Optimum gibt. Dahingegen haben L_c und n nur Bereiche, in denen der Parameter liegen sollte, so muss L_c relativ klein sein, da es keine guten Lösungen mehr mit L_c größer als 4.5 gibt. Die besten Lösungen für n liegen im Bereich von 1.5

bis 3.5, der Abfall der Bewegungsgeschwindigkeit von $v_1 + v_2$ auf v_1 sollte also moderat erfolgen. Für v_2 kann man eigentlich keinen optimalen Parameterwert angeben, da über den gesamten verwendeten Bereich gute Lösungen gefunden wurden. Die Parameter v_2 , L_c und n können sich innerhalb gewisser Grenzen gegenseitig kompensieren während für v_1 ein eindeutiger optimaler Wert gefunden werden kann.

Tabelle 20 gibt die besten gefundenen Parameter wieder. Sie werden in allen weiteren Rechnungen verwendet. In Abbildung 39 ist die Bewegungsgeschwindigkeit im inneren des Proteasoms in Abhängigkeit von der Anzahl der Reste jenseits des aktiven Zentrums dargestellt. Die Bewegungsgeschwindigkeit ist auf die durchschnittliche proteasomale Schnittrate normiert. Es ist zu erkennen, dass die Bewegungsgeschwindigkeit von einer sehr hohen initialen Rate abfällt. Für diese Verminderung der Vorwärtsrate genügen schon wenige Aminosäuren jenseits des aktiven Zentrums aus. Die Bewegungsrate ist immer um ein vielfaches höher als die durchschnittliche Schnittrate des Proteasoms und auch bei sehr günstigen Schnittmotiven liegen Bewegungsrate und Schnittrate in der gleichen Größenordnung.

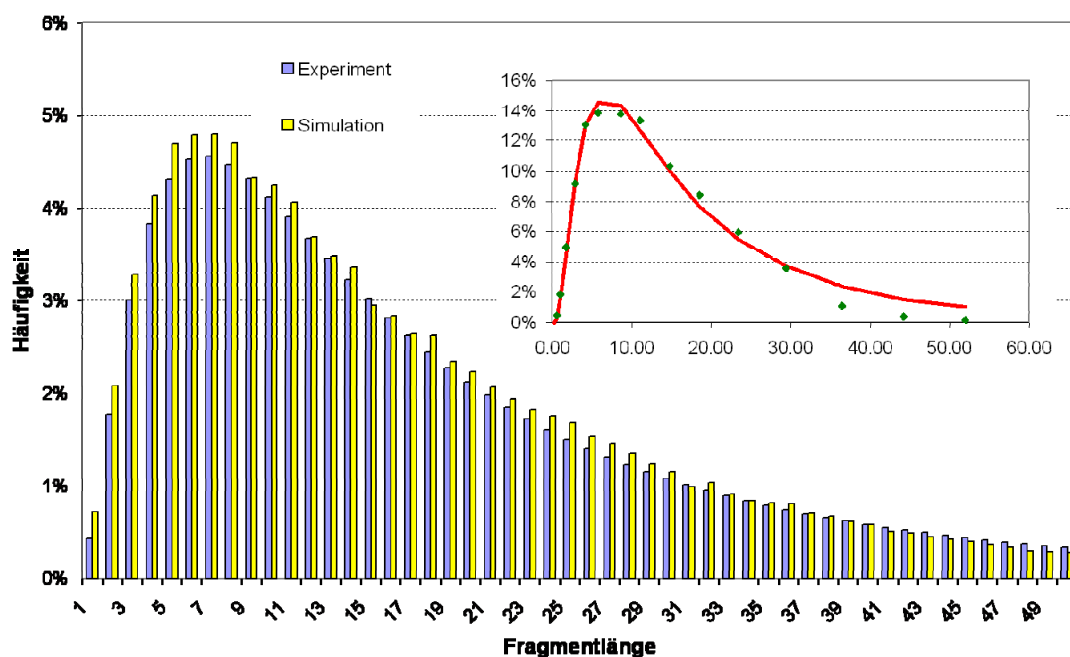


Abbildung 37: Fragmentlängen sind lognormal verteilt

Im Inset sind die experimentellen Datenpunkte (grün) und die daran angepasste Längenverteilung dargestellt. Im Hauptfenster sind die aus dieser Längenverteilung interpolierten Werte für jede Fragmentlänge als Experiment (blau) dargestellt. Die gelben Balken geben die Längenverteilung der Simulation mit dem geringsten Abstand zum Experiment an. Im Experiment wurde Ovalbumin verdaut. Die y-Achsen der beiden Graphen haben unterschiedliche Werte, da die Klassenbreiten nicht identisch sind.

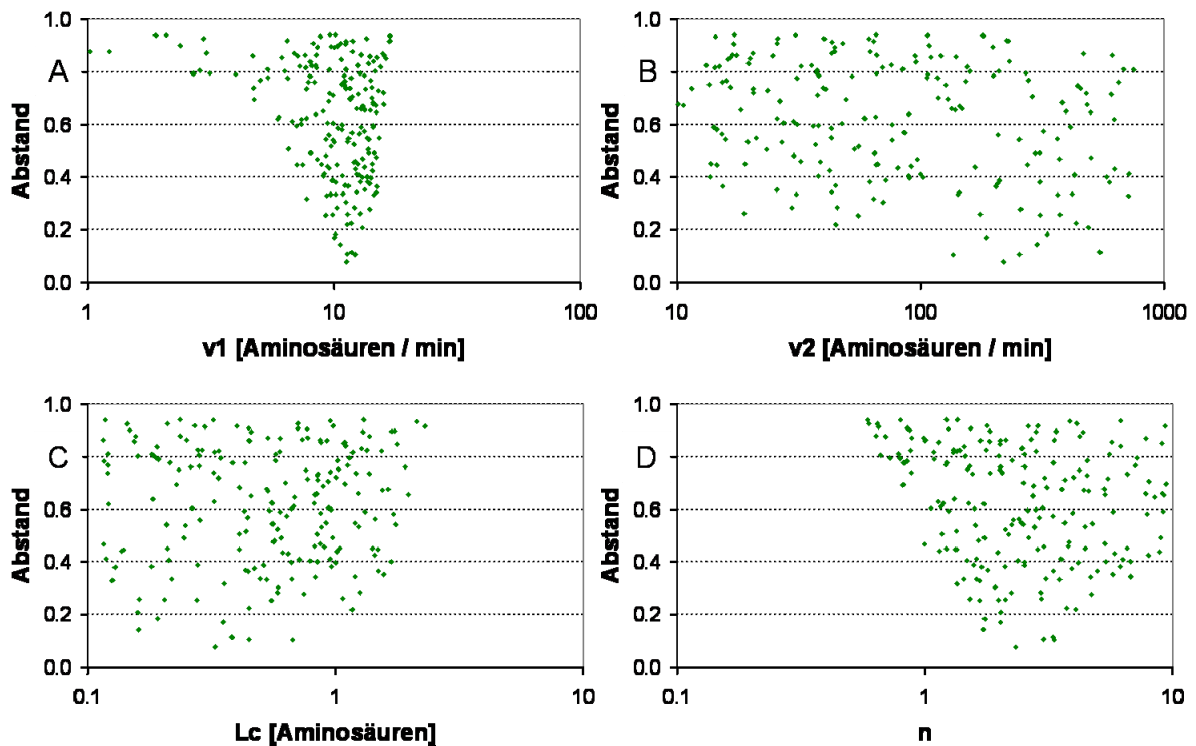


Abbildung 38: Abstand zwischen experimenteller und simulierter Längenverteilung

Die Diagramme zeigen den Abstand zwischen simulierter und experimenteller Längenverteilung gegen die vier verwendeten Parameter (Abbildung A: v1; B: v2; C: L_c; D: n). Die Abbildungen sind auf die 220 besten Parametertupel begrenzt. Die Parameter v1 und v2 sind auf eine Basisrate des Proteasoms von 1 normiert.

Tabelle 20: Optimale Parameter für Längenverteilung

Parameter	Wert	Dimension
v1	11.3	aa / min
v2	440	aa / min
Lc	0.656	aa
n	2.34	

Translation der Peptidkette in Aminosäureeinheiten pro Minute. Die Werte für v1 und v2 sind auf eine Basisrate des Proteasoms von 1 normiert.

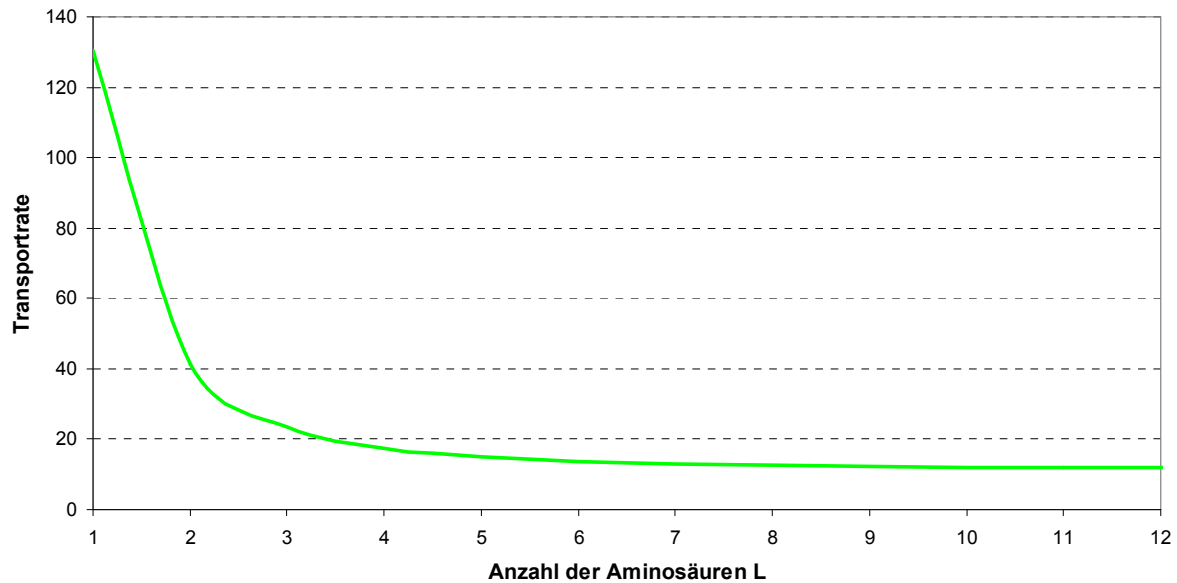


Abbildung 39: Transportrate der Peptide im Proteasom

Die Transportrate der Peptide im Proteasom ist gegen die Anzahl der Aminosäuren jenseits des aktiven Zentrums dargestellt. Die Werte sind auf eine Basisrate des Proteasoms von 1 normiert. Die Maximalrate des Proteasoms liegt in der Größenordnung von 10. Somit hat selbst die günstigste Aminosäuresequenz nur eine Chance von 50 % geschnitten zu werden und auch das nur, wenn bereits genügend Aminosäuren jenseits des aktiven Zentrums sind.

4.2.2 Fragmente lassen sich nicht gut vorhersagen

Beim proteasomalen Verdau der Enolase konnten experimentell 80 Fragmente identifiziert werden[130]. Dieses Experiment wurde mit den zuvor bestimmten Parametern simuliert. Alle vom Proteasom erhaltenen Fragmente wurden nach Häufigkeit sortiert. Die Häufigkeit wird als Cutoff Wert für die experimentelle Detektierbarkeit festgelegt und systematisch variiert, wodurch eine zu den in Abschnitt 3 verwendeten ROC Kurven vergleichbare Abbildung erstellt werden kann. Bei der klassischen Darstellung von Sensitivität gegen 1-Spezifität (Abbildung 40 c) würde man zu der falschen Annahme kommen, dass die Vorhersage sehr gut funktioniert, da der AUC Wert 0.92 beträgt. Dieses ist aber dadurch bedingt, dass sehr viele Fragmente theoretisch erzeugbar wären ($95 \cdot 266 = 436 \cdot 437 / 2$; die Enolase hat 436 Aminosäuren), von denen jedoch sehr viele aufgrund der eingestellten Längenverteilung in der Simulation nicht produziert werden.

Es müssen 2000 bis 2500 Fragmente als vom Proteasom erzeugbar klassifiziert werden, um 50 der 80 experimentell gefundenen Fragmente richtig klassifiziert zu haben (Abbildung 40 a). Dieses ist etwas besser, als wenn man zufällig Fragmente aus der oben be-

schriebenen Längenverteilung ziehen würde und wenn verlangt wird, dass alle 80 Fragmente gefunden werden, ist die Vorhersage nur noch marginal besser als die Zufallsvorhersage (Abbildung 40 b).

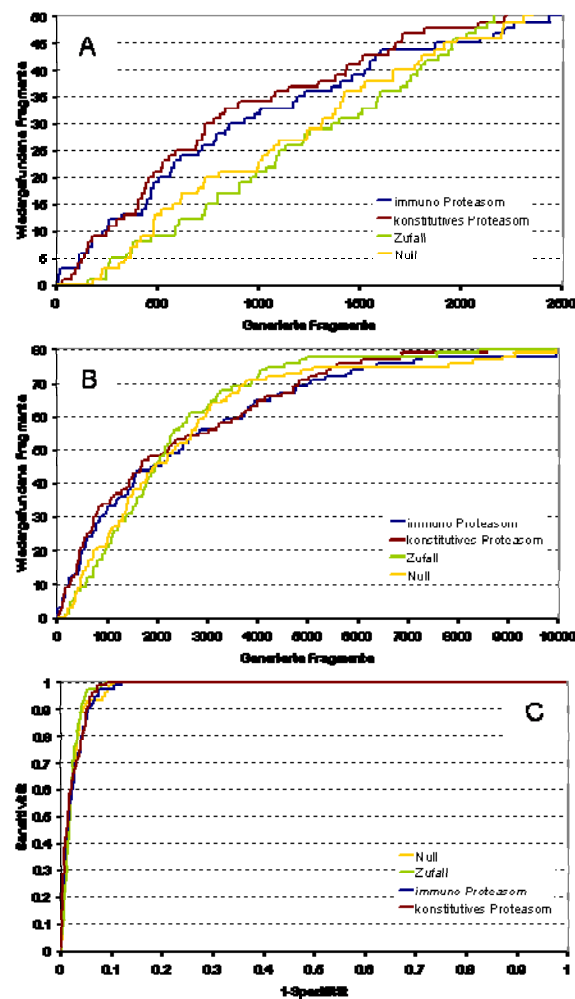


Abbildung 40: Performance der Fragmenterstellung

In den Abbildungen A und B sind jeweils die Anzahl der richtig als vom Proteasom hergestellt klassifizierten Fragmente gegen die Gesamtzahl der als vom Proteasom herstellbar klassifizierten Fragmente dargestellt. Die Kurve „Zufall“ wurde erstellt, indem zufällig Fragmente entsprechend der Längenverteilung aus dem Protein genommen und als hergestelltes Fragment definiert wurden. Für die Kurve „Null“ wurden die Einträge in der Vorhersagematrix für die proteasomale Schnittwahrscheinlichkeit auf Null gesetzt. Somit hat das Proteasom an jeder Position die Basisrate als Schnittrate. In Abbildung C ist die klassische ROC Kurve als Darstellung der Ergebnisse gewählt. Die gute Performance ($AUC = 0.92$) wird aber nur durch die Verwendung einer geeigneten Längenverteilung erreicht.

Festzuhalten bleibt noch, dass sich weder bei der Längenverteilung noch bei der Fragmentvorhersage etwas ändert, wenn man die Richtung, in der das Protein das Proteasom durchquert ändert.

4.2.3 Einbeziehen von Sekundärstrukturinformationen verbessert die Vorhersage nicht

Ein Versuch, diese unbefriedigenden Ergebnisse der Fragmentvorhersage zu verbessern bestand darin, Sekundärstrukturinformationen über das Protein mit in die Vorhersage einzubeziehen. Die Hypothese wäre hier, dass das Protein nicht in komplett gestreckter Konformation im inneren des Proteasoms vorliegt, sondern Sekundärstrukturelemente oder vergleichbare Strukturen bildet.

Im dieses zu Testen wurde die Sekundärstrukturinformation über die Enolase aus der ExPASy Datenbank[167, 168] verwendet. Es wurden drei neue Parameter implementiert, die die Schnittrate des Proteasoms modifizieren, in Abhängigkeit davon, ob das Substrat am aktiven Zentrum als α -Helix, β -Faltblatt oder als Coil vorliegt.

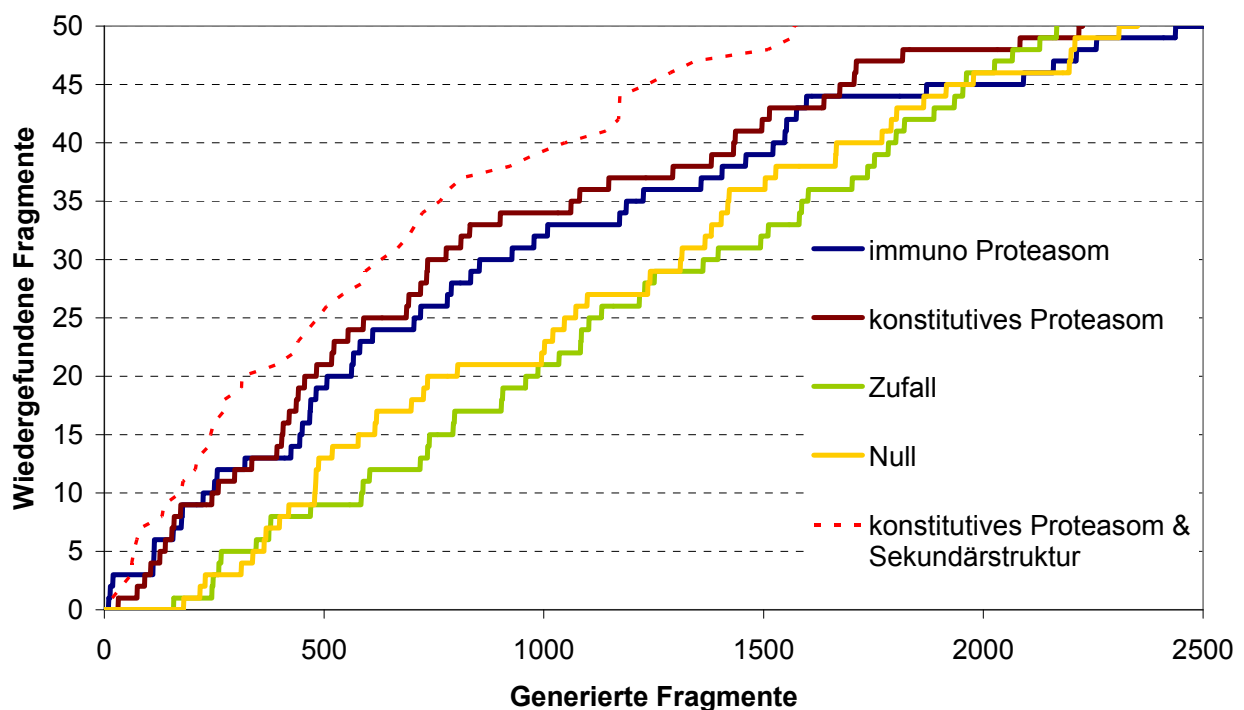


Abbildung 41: Güte der Fragmenterstellung mit Sekundärstrukturinformation

Die Abbildung ist wie Abbildung 40 A. Die hier erscheinende Verbesserung der Vorhersage ist auf die Anpassung der Längenverteilung auf den Verdau der Enolase zurückzuführen.

Im ersten Szenario wurde die gesamte Schnittrate des Proteasoms geändert, je nachdem welche Konformation die Aminosäuren an P1 und P1' aufweisen.

Es konnte keine Verbesserung der Vorhersage erzielt werden. Durch optimales Wählen der Parameter verbessert sich die Vorhersage scheinbar (Abbildung 41). Dieses ist jedoch auf eine generelle Erhöhung der proteasomalen Schnittrate zurückzuführen, dass heißt, es wird die gleiche Wirkung erzielt, wenn an jeder Stelle unabhängig von der Sekundärstruktur die Wahrscheinlichkeit erhöht wird zu schneiden oder äquivalent die Bewegungsraten v_1 und v_2 vermindert werden.

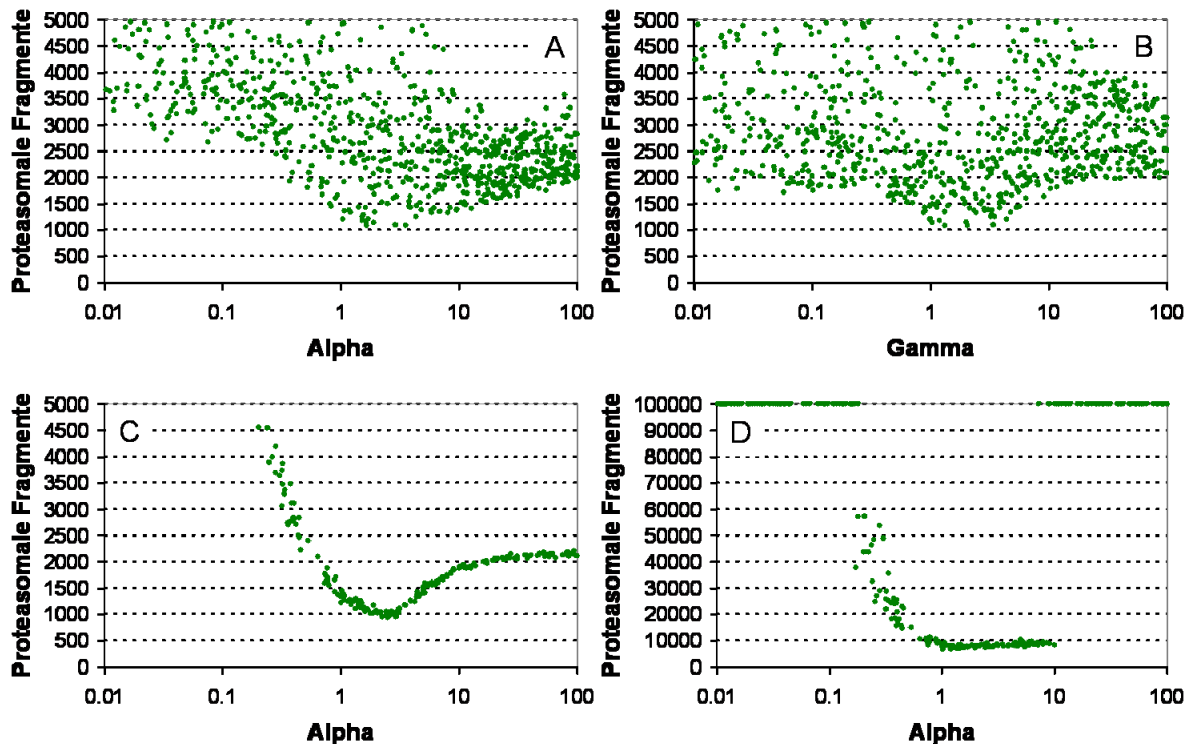


Abbildung 42: Parameterplot für Sekundärstrukturinformation

In den vier Abbildungen sind jeweils die als vom Proteasom herstellbar zu klassifizierenden Fragmente dargestellt, um eine bestimmte Anzahl (A bis C: 40; D: 80) von tatsächlich gebildeten Fragmenten zu erhalten. In A ist der Modifikator für die α -Helix Regionen auf der x-Achse variiert. In B der Modifikator für Coil Regionen. In C und D wurde der Modifikator für alle Regionen gleich gewählt und auf der x-Achse dargestellt. Die Abhängigkeiten in A und B sind mit der in C zu beobachtenden Abhängigkeit zu erklären. Wenn die Modifikatoren außerhalb eines geeigneten Bereichs gewählt werden können nicht mehr alle Fragmente während einer Simulation gebildet werden (Datenpunkte am oberen Rand; Abbildung D).

Die verwendeten Parameter der Längenverteilung waren nicht an dieses spezielle Experiment angepasst und durch eine entsprechende Optimierung lässt sich die Vorhersage verbessern, was aber nicht auf die Sekundärstruktur der Enolase zurückzuführen ist.

Die Längenverteilungen der gebildeten Fragmente zweier Proteine unterscheiden sich schon daher, dass jedes Protein eine eigene Aufnahmerate in das Proteasom hat und bei Langzeitverdaus eine unterschiedlich lange Zeit vergeht, bis 50 Prozent des Substrats verdaut ist. In dieser Zeit können gebildete Fragmente weiter verkürzt werden und somit die Längenverteilung verändern.

Ausgewählte Parameterplots sind in Abbildung 42 dargestellt.

Zusätzlich wurden diese Rechnungen in einem zweiten Szenario durchgeführt. Hier wurde jeweils der Beitrag zur proteasomale Schnittrate der einzelnen Positionen im beitragsenden Sequenzfenster (P6 bis P4') modifiziert. Auch in diesen Rechnungen konnte keine Verbesserung der Vorhersagequalität der Fragmente erreicht werden.

4.3 Diskussion

Zusammenfassend bleibt festzuhalten, dass man zwar mit der Vorhersage von proteasomalen Schnittraten in der Lage ist, die präsentierten Antigene mit moderater Präzision vorherzusagen aber es ist nicht zufrieden stellend möglich Fragmente vorherzusagen.

An dieser Stelle muss man sich natürlich die Frage stellen, ob das verwendete Modell zu einfach oder gar falsch ist. Hierzu ist anzuführen, dass zwei mittlerweile publizierte Modelle [165, 169] keine besseren Ergebnisse liefern. Das Modell von Mishto et al. macht die Annahme, dass das Substrat in das Proteasom aufgenommen wird, dort geschnitten wird und letztendlich Fragmente wieder freigesetzt werden. Im Proteasom gebildete Fragmente, die noch nicht freigegeben wurden können weiter in kürzere Peptide hydrolysiert werden. Das System wird in diesem Modell als Mastergleichung dargestellt und alle Prozesse (Aufnahme, Abgabe und Schneiden) sind längenabhängig. In beiden publizierten Arbeiten wird festgestellt, dass man die spezifischen positionsabhängigen Schnittpräferenzen des Proteasoms benötigt, um die richtigen Fragmente zu erzeugen. Diese Forderung macht für eine Prädiktion von Fragmenten aus Proteinen natürlich keinen Sinn. Damit bleibt für das hier vorgestellte Modell festzuhalten, dass die mangelnde Performance nicht auf die Struktur des Modells zurückzuführen ist, sondern dass das Modell zu einfach ist, also Prozesse, die das Schnittverhalten des Proteasoms bestimmen, nicht beinhaltet. Das Problem ist jedoch, dass hierfür keine experimentellen Daten zur Verfügung stehen.

Das dem Modell der Fragmentbildung zugrunde liegende Modell der Schnittstellenvorhersage beruht auf der Annahme des unabhängigen Beitrags der Aminosäuren in der Nähe des aktiven Zentrums. Wenn das Schnittverhalten aber durch kooperative Effekte der Aminosäuren determiniert wird, kann es nur in begrenztem Maße mit diesem Modell vorhergesagt werden. Eine Hypothese, dass die Peptide lokale Sekundärstrukturen bilden, die das Schnittverhalten beeinflussen, konnte hier falsifiziert werden. Es ist dabei aber nicht auszuschließen, dass andere Effekte das Schnittverhalten bestimmen. Um dieses adäquat in einem Modell zu berücksichtigen bräuchte man eine Parametrisierung, die ähnlich den Paarkoeffizienten wäre. Dafür stehen aber bei weitem zu wenig experimentelle Daten zu Proteinverdaus zur Verfügung.

Weiterhin wurde in verschiedenen Arbeiten gezeigt, dass Peptide das Schnittverhalten des Proteasoms verändern können[170, 171, 172]. Dieses kann von dem hier vorgestellten Modell nicht erfasst werden und bildet ebenfalls einen möglichen Grund für das schlechte Abschneiden des Modells bei der Fragmenterstellung. *In vivo* ist die Situation noch komplizierter, da hier auch noch die modulierenden Untereinheiten das Schnittverhalten des Proteasoms verändern können[173].

Die mit dem Modell gemachten Vorhersagen sind immerhin etwas besser als zufälliges Auswählen der Fragmente (Abbildung 40 a). Weiterhin ist anzunehmen, dass in den sehr aufwändigen Experimenten nicht alle gebildeten Fragmente gefunden werden und sehr wohl deutlich mehr verschiedene Fragmente vom Proteasom gebildet werden können.

Es sind keine Informationen über den Mechanismus des Schneidens und den Transport der Peptide im inneren des Proteasoms verfügbar. Daher sind alle Modelle, die qualitativ gleichwertige Resultate liefern gleichberechtigt und Modelle, die weniger Annahmen machen beziehungsweise weniger Parameter beinhalten sind meiner Meinung nach zu bevorzugen. Daher wird dieses Modell in Abschnitt 5.2 verwendet, wohl wissend, dass deutlich mehr proteasomale Fragmente erstellt werden, als im Experiment beobachtet werden.

An der Erforschung der proteasomalen Fragmentbildung werden in Zukunft noch viele Anstrengungen zu erbringen sein.

5 Kinetische Modellierung der Epitop Generierung

In diesem Abschnitt sollen die einzelnen in Kapitel 3 und 4 dargestellten Modelle der Teilschritte der Antigenpräsentation zu kinetischen Modellen zusammengefügt werden. Die Teilmodelle wurden zu verschiedenen Zeiten entwickelt und standen zum Zeitpunkt der Durchführung dieser Rechnungen nicht in ihrer Gesamtheit zur Verfügung.

Wie die einzelnen Antigene generiert und prozessiert werden, ist aus molekulardynamischer Sicht weitgehend unbekannt. Die hier verwendeten Gleichungen zum Beschreiben der Antigenpräsentation sind deshalb sehr einfach gehalten. Das Ziel dieser Modellierung ist die Klärung, ob die bekannten Prozesse ausreichend für die Beschreibung des beobachteten Systems sind und ob wir durch die Modellierung ein besseres Verständnis des Prozesses entwickeln können. Eine Verbesserung der Vorhersagequalität der Modelle ist kein Ziel dieses Abschnitts.

5.1 Antigenprozessierung aus pp89-25mer

Das pp89-25mer Oligopeptid¹⁰ wurde in der Arbeitsgruppe von Professor Kloetzel am Institut für Biochemie der Charité sehr gut charakterisiert. So ist bekannt, welche Fragmente mit welcher Rate vom Proteasom beim Abbau gebildet werden[129] und auch, welcher Teil als Antigen an der Zelloberfläche präsentiert wird und dass N-terminal verlängerte Vorläufer des präsentierten Epitops in das ER transportiert werden[174]. Das Protein pp89 ist ein Protein des murinen Cytomegalovirus.

5.1.1 Beschreibung des verwendeten Modells

Das 25mer des viralen Proteins pp89 wird mit konstanter Rate verdaut und da die Bildungsraten der Fragmente bestimmt worden sind, werden somit die einzelnen Fragmente direkt erzeugt. Dieses ist die direkte Wiedergabe des gemessenen proteasomalen Abbaus und es bedarf somit keines Modells für diesen ersten Schritt des Pathways. Fragmente im Cytosol können entweder mittels TAP in das ER transportiert werden oder werden durch cytosolische Aminopeptidasen N-terminal um eine Aminosäure verkürzt (Abbildung 43).

Die Simulation wird stochastisch als Monte Carlo Simulation durchgeführt, wobei der Gillespie Algorithmus[166] verwendet wird. Die Erzeugung eines Fragments, der Transport in das ER eines Fragments sowie das cytosolische Trimmen eines Peptids stellen jeweils ein Elementarereignis dar. Die Raten für die proteasomale Erzeugung der Fragmente sind

¹⁰ Sequenz: RLMYDMYPHFMPNTLGPSEKRVWMS

experimentell bekannt. Es werden jeweils entweder 20 Proteine pro Minute abgebaut oder 200. Diese Werte sind physiologisch für Proteine mit geringer bis mittlerer Konzentration im Cytosol[30]. Es wurde keine Änderung in der Menge des verfügbaren Substrats betrachtet. Für die Transportraten der Fragmente in das ER wird das entwickelte Modell[87] verwendet. Das Modell liefert jedoch nur relative IC50 Werte beziehungsweise Ratenkonstanten. Die Rate für das Bezugspeptid muss hierbei noch festgelegt werden und wurde so gewählt, dass der Transportprozess eine Halbwertszeit von 120 Sekunden hat. Die Raten der Aminopeptidasen waren zum Zeitpunkt dieser Simulationen noch nicht gemessen und es wurden verschiedene Werte (Tabelle 21) verwendet, um den potentiellen Einfluss der Aminopeptidasen auf die Antigenpräsentation abzuschätzen. Als Basiswert, um den die einzelnen Raten für die individuellen Aminosäuren streuen, wurden 5 Sekunden gewählt, was in Übereinstimmung mit experimentellen Befunden ist[88]. Weder TAP Transport noch N-terminales Trimmen können in den hier durchgeführten Rechnungen gesättigt werden.

Der Gillespie Algorithmus erlaubt die Simulation des Systems, indem zu einem Zustand alle möglichen Prozesse bestimmt werden, von denen dann, entsprechend ihrer relativen Rate, einer zufällig ausgewählt und ausgeführt wird. Das System befindet sich dann in einem neuen Zustand. Hierbei kann zusätzlich noch eine Systemzeit mitgeführt werden, indem die für jeden Systemschritt vergehende Zeit zufällig aus einer logarithmischen Verteilung gezogen wird und mit der zu jedem Systemzustand gehörenden Gesamtrate (=Summe aller Raten aller Elementarprozesse) normiert wird. Eine Simulation ist somit ein Pfad durch den Zustandsraum des Systems und indem ausreichend viele Simulationen durchgeführt werden erhält man eine gute Approximation der Lösung der zugehörigen Mastergleichung ohne das System deterministisch lösen zu müssen. In dem hier gerechneten relativ kleinen Beispiel wäre die deterministische Lösung durchaus noch bestimmbar, jedoch für große Proteine kommt man relativ schnell an die Grenzen der vertretbaren Rechenzeit und Speicherbedarfs.

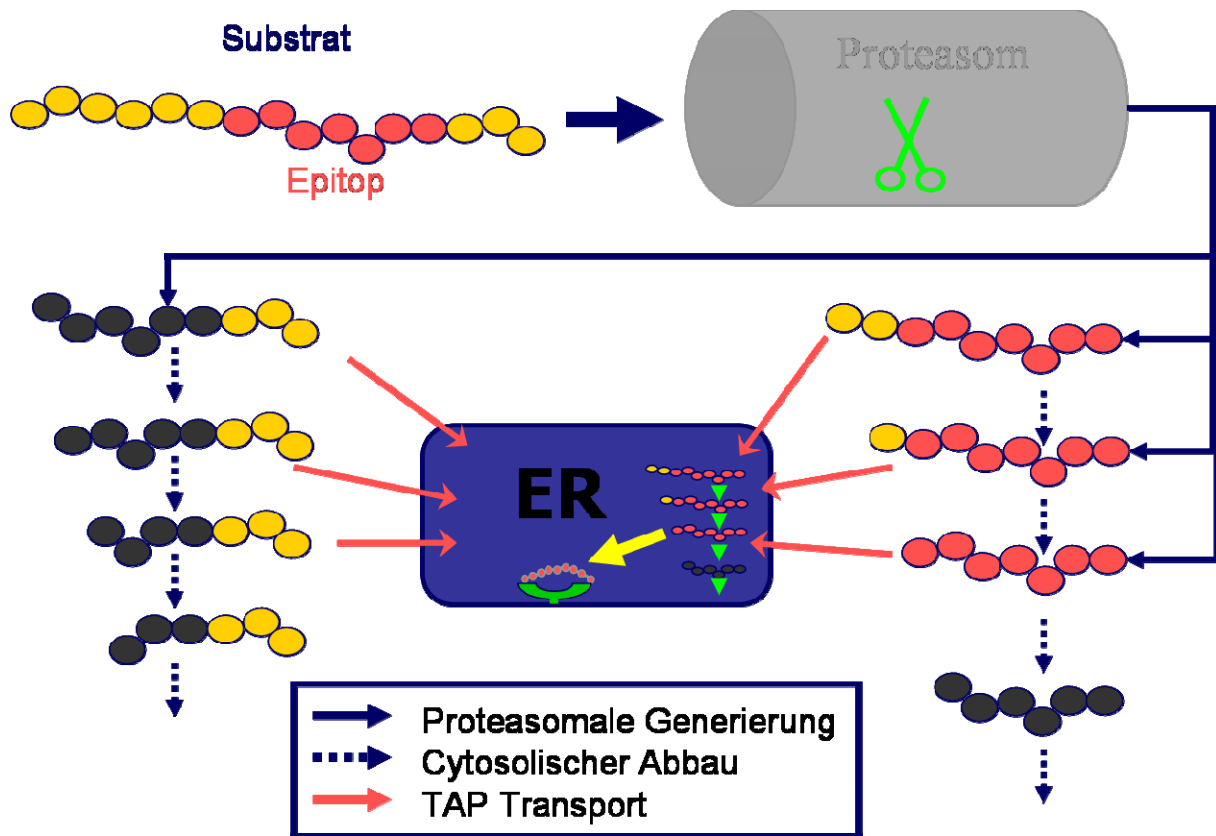


Abbildung 43: Schematische Übersicht des kinetischen Modells

Das Substrat wird vom Proteasom in kürzere Fragmente gespalten. Diese Fragmente werden entweder im Cytosol um jeweils eine N-terminale Aminosäure verkürzt oder in das ER transportiert. Das im Substrat enthaltene Epitop kann durch das Proteasom aber auch durch Aminopeptidasen zerstört werden. Als Precursoren gelten nur Fragmente, die das intakte Epitop am C-Terminus enthalten. Der Transport der Fragmente durch TAP in das ER ist nur möglich, wenn das Peptid zwischen Minimal- und Maximallänge für diesen Prozess ist. Die Prozesse im ER, das heißt der Abbau durch endoplasmatische Aminopeptidasen (grüne Pfeile) sowie die Bindung des Epitops an einen MHC Komplex (gelber Pfeil) werden im Kapitel 5.2 simuliert.

Tabelle 21: Ratenkonstanten für den cytosolischen Abbau der Peptide

	Set 1 gleich	Set 2 Zufall 1	Set 3 Zufall 2	Set 4 Zufall 3	Set 5 Epitop	Set 6 Precursor
A	8.00	175.54	6.68	6.91	8.00	8.00
C	8.00	64.45	0.07	429.65	8.00	8.00
D	8.00	0.12	76.80	0.36	40.00	0.20
E	8.00	0.05	1.16	0.98	8.00	8.00
F	8.00	0.17	1.76	0.58	8.00	8.00
G	8.00	31.53	69.77	2.26	8.00	8.00
H	8.00	11.59	3.93	1.72	8.00	8.00
I	8.00	0.31	0.48	0.13	8.00	8.00
K	8.00	14.67	0.66	0.59	8.00	8.00
L	8.00	0.28	2.20	0.48	40.00	0.20
M	8.00	0.24	0.17	0.02	40.00	0.20
N	8.00	0.04	3.82	0.80	8.00	8.00
P	8.00	0.35	1.15	4.08	8.00	8.00
Q	8.00	0.01	0.01	75.63	8.00	8.00
R	8.00	0.21	9.11	36.49	8.00	8.00
S	8.00	0.30	5.02	0.31	8.00	8.00
T	8.00	37.72	71.57	0.22	8.00	8.00
V	8.00	4.49	21.51	0.51	8.00	8.00
W	8.00	0.55	0.10	3.00	8.00	8.00
Y	8.00	0.37	0.56	6.63	0.01	800.00

Die Ratenkonstanten sind so gewählt, dass die Peptide eine Lebenszeit von 5 Sekunden haben. Die Werte der Sets 2, 3 und 4 wurden zufällig einer Lognormalverteilung entnommen. In den Sets 5 und 6 wurden die Raten für einzelne Aminosäuren so angepasst, dass das Epitop oder seine Precursoren maximal stabil beziehungsweise instabil sind.

5.1.2 Ergebnisse der Simulation

Einfluss der Aminopeptidaseraten auf die Antigenpräsentation

Es wurden Simulationen mit allen Sets aus Tabelle 21 durchgeführt. Nach einer Startphase stellt sich im System ein stationärer Zustand ein, bei dem das ER mit konstanter Rate aller Fragmente beliefert wird. Diese Raten des Transports der Peptide in das ER wurden gegeneinander aufgetragen (Abbildung 44). Es zeigt sich, dass die Raten der Aminopeptidaseaktivitäten auf die einzelnen N-terminalen Aminosäuren einen Einfluss auf die Antigenpräsentation haben können. So schwanken die Fragmentmengen, die in das ER transportiert werden in den hier durchgeführten Simulationen um ein bis zwei Größenordnungen (Abbildung 44 (a-c)). Dieses ist ein deutlich geringerer Einfluss, als die Selektivitäten von TAP oder gar der MHC Moleküle bewirken, jedoch kann auch eine solche geringe Modifikation entscheiden, ob ein potentiell Antigen in ausreichend großen Stückzahlen präsentiert wird. Die in Abschnitt 3.4 dargestellten Untersuchungen haben

gezeigt, dass der Einfluss des N-terminalen Trimmens durchaus zu beobachten ist. Durch den Transport von Precursoren des Epitops in das ER kann der Einfluss der Selektivität der Aminopeptidasen etwas verringert werden, da meist nicht alle Precursoren eines Epitops gleichermaßen abgebaut werden. Dieses ist in allen vier Teilen von Abbildung 44 zu erkennen.

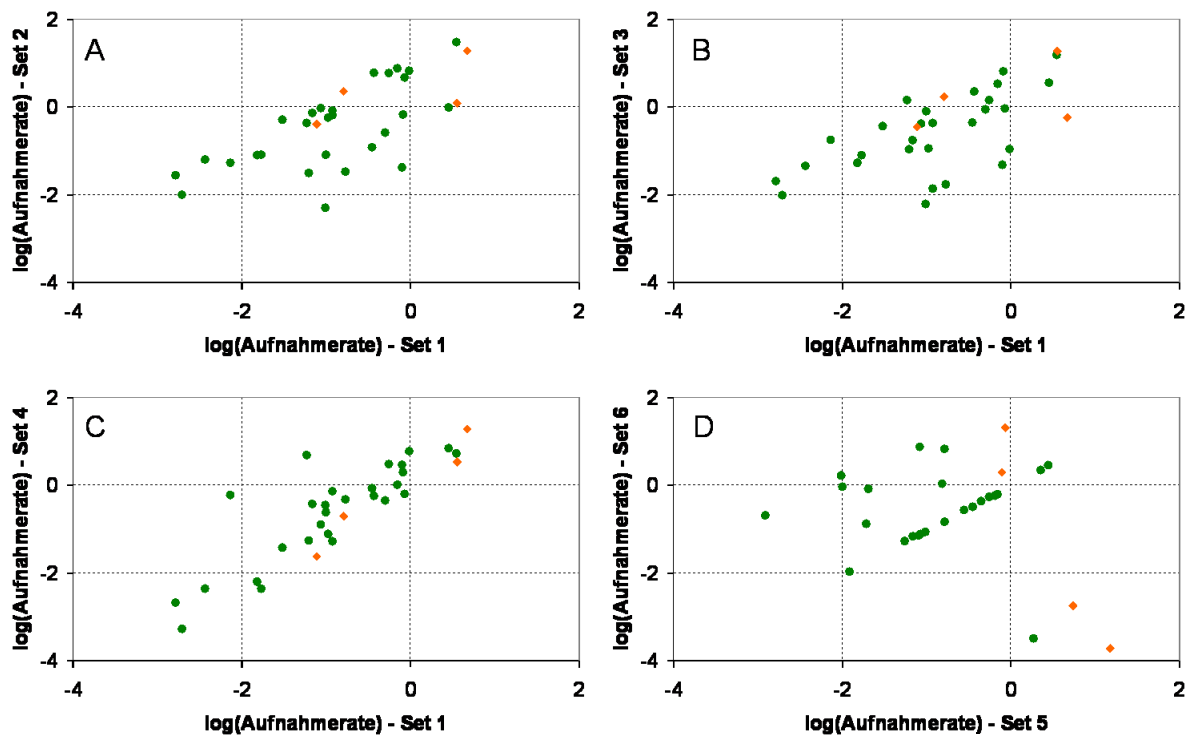


Abbildung 44: Varianz der Aufnahmezeiten der Peptide in das ER abhängig vom cytosolischen Trimmen

Die Abbildung gibt die Variabilität der Aufnahmezeiten der Peptide in das ER für die verschiedenen Sets in Tabelle 21 wieder. Das Epitop und seine Precursoren sind in rot dargestellt.

In den weiteren Rechnungen zum 25mer des pp89 wurde Set 1 verwendet, da keine besseren experimentellen Daten zur Verfügung standen.

Genauere Untersuchung des präsentierten Epitops und seiner Precursoren

Zur Untersuchung des Einflusses der cytosolischen Prozesse auf die Antigenpräsentation wurde das System 1000-mal für 30 Minuten Systemzeit simuliert. Hierbei wurden das Epitop sowie seine Precursoren genauer betrachtet.

Im jetzt folgenden Abschnitt wird das präsentierte Epitop als 9mer bezeichnet und die drei Precursoren entsprechend als 10mer, 11mer und 12mer. Wie experimentell be-

stimmt, wird vom Proteasom hauptsächlich das 11mer hergestellt. Das 12mer und das Epitop werden jeweils mit etwa einem Drittel dieser Effizienz produziert und die Rate des 10mers ist vernachlässigbar gering (Abbildung 45 a). Wenn man hieraufhin jedoch die in das ER transportierten Peptide (Abbildung 45 b) genauer betrachtet, fällt auf, dass das 10mer mit fast der gleichen Rate wie das 11mer transportiert wird. Das 12mer und das eigentliche Epitop gelangen mit nur sehr geringen Stückzahlen in das ER.

Die Anzahl der in das ER transportierten Epitope ohne Precursoren würde nicht ausreichen, um ausreichend Antigen für eine Immunantwort an der Zelloberfläche präsentieren zu können. In fast 10 Prozent aller Simulationen wurde nicht ein Epitop in das ER transportiert (Abbildung 47 a). In allen Simulationen wurden aber mehr als 200 Precursoren und Epitope in das ER transportiert (Abbildung 47), was bei einer angenommenen Effizienz des weiteren Präsentationswegs von 10 Prozent ausreichend für die Präsentation und darauf folgende Immunantwort wäre. Der Hauptteil wird hierbei von 11mer und 10mer getragen.

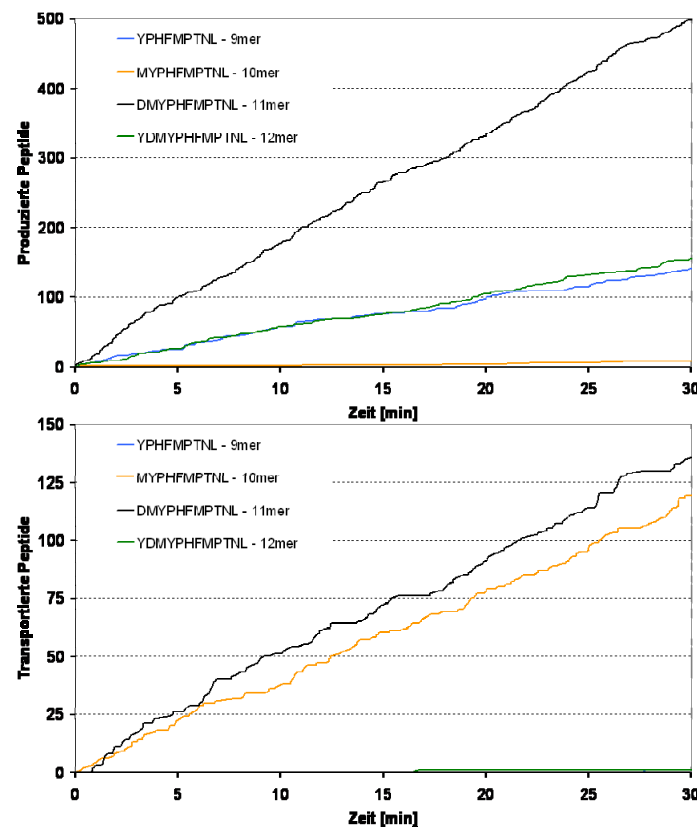


Abbildung 45: Zeitverläufe vom Epitop und seinen Precursoren

Die Abbildung gibt die zeitlichen Verläufe für das Epitop und seine Precursoren wieder. Oben ist die vom Proteasom erstellte Menge der einzelnen Peptide dargestellt, unten die Anzahl der in das ER transportierten Peptide.

Abbildung 46 zeigt die Beladungsverteilung des Cytosols mit den einzelnen Peptiden. Diese Statistik wurde zu verschiedenen Zeitpunkten der 1000 Simulationen erstellt. Hier zeigt sich, dass das 12mer am wenigsten präsent ist (durchschnittlich 0.7 Epitope zu einem Zeitpunkt pro Zelle), während das Epitop, das proteasomal mit annähernd der gleichen Rate hergestellt wird, am häufigsten vertreten ist (2.3 Epitope durchschnittlich). Das 11mer und auch das 10mer weisen mit 2.1 beziehungsweise 1.7 Epitopen pro Zelle nur etwas weniger Abundanz auf als das eigentliche Epitop. Für das 10mer ist dieses besonders bemerkenswert, da es vom Proteasom fast nicht hergestellt wird.

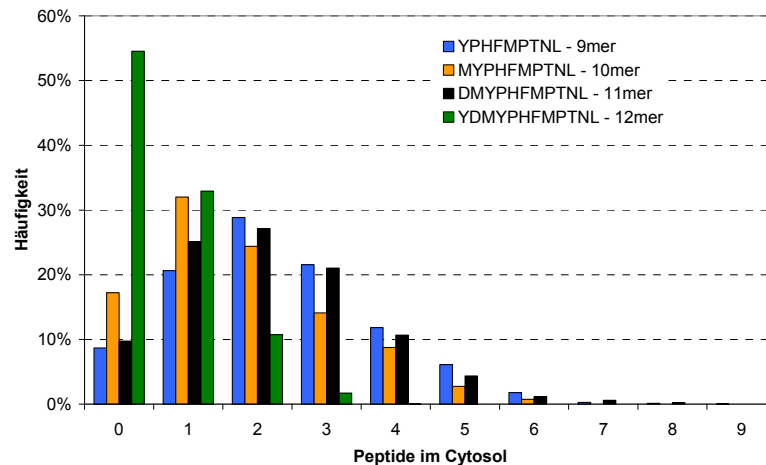


Abbildung 46: Anzahl von Epitop und Precursor Molekülen im Cytosol

Die Anzahl der einzelnen Peptide im Cytosol zu verschiedenen Zeitpunkten. Die Anzahl der Peptide im Cytosol ist immer gering und zu einzelnen Zeitpunkten ist kein einziges Peptid eines Precursors zu finden.

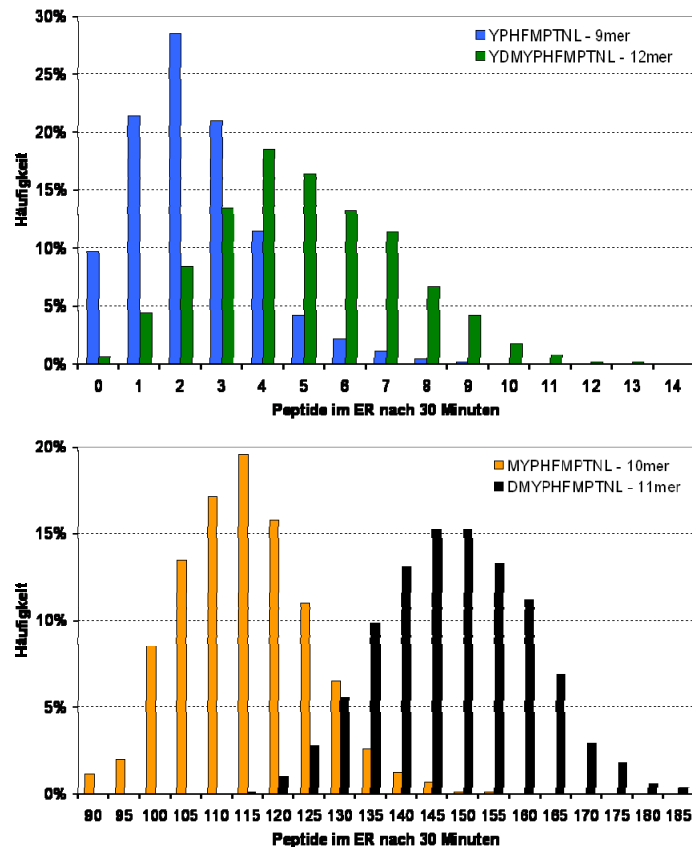


Abbildung 47: Verteilung der Peptidmengen im ER nach 30 Minuten

Die Abbildung gibt die Verteilung der Peptidmengen des Epitops und seiner Precursoren im ER nach 1000 Simulationen mit jeweils 30 Minuten wieder. Oben sind die Peptide mit geringen Mengen, unten die Peptide mit hohen Mengen. Es ist bemerkenswert, dass in 98 der 1000 Simulationen nicht ein einziges Epitop in das ER transportiert wurde.

5.1.3 Diskussion

Precursoren stellen einen Schutz der präsentierbaren Epitope vor dem Abbau durch Aminopeptidasen dar. Aber sie erweitern ebenfalls das Spektrum der von TAP transportierbaren Peptide und stellen so sicher, dass ausreichende Mengen an Epitop zur Verfügung stehen. Das 11mer wird hauptsächlich produziert und experimentell wurde gezeigt, dass es auch als Precursor in das ER transportiert wird [174]. Die hier durchgeführten Simulationen legen nahe, dass das 10mer eine genauso wichtige Rolle spielt, da es mit fast der gleichen Rate in das ER transportiert wird. Dieses scheint zuerst einen Widerspruch darzustellen, da mehr 10mere in das ER transportiert werden, als vom Proteasom hergestellt werden. Die cytosolischen Aminopeptidasen bauen die längeren Precursoren ab und stellen dabei kürzere Peptide her, die dann in das ER transportiert werden können. So ist die

Menge aller 11mer Vorläufer, die in das ER transportiert werden können, die Summe der im Cytosol abgebauten 12mere und der vom Proteasom hergestellten 11mere. Für das 10mer verhält es sich ähnlich, nur ist die proteasomale Generierung gering und alle abgebauten 11mere, also auch die, die schon aus 12meren gebildet wurden, tragen zur Menge der prozessierbaren 10mere bei. Diese Art der Betrachtung erklärt auch, warum das Epitop die höchste durchschnittliche Beladung der vier betrachteten cytosolischen Peptide aufweist, obwohl es nur mit mittlerer Effizienz vom Proteasom gebildet wird. Die Effizienz des Transports durch TAP entscheidet dann, welches Peptid in das ER gelangt. Durch den Abbau der Peptide im Cytosol werden zu den vom Proteasom erzeugten noch weitere Peptide erzeugt, die dann von TAP transportiert werden können. So wird es ermöglicht, dass auch Peptide präsentiert werden können, die mit geringer Effizienz vom Proteasom erzeugt und eine geringe TAP Bindung aufweisen. Es muss jedoch mindestens ein Precursor proteasomal erzeugt werden, und mindestens ein durch TAP transportierbares Peptid muss während des cytosolischen Abbaus der Precursoren erzeugt werden. Jedoch ist zu beachten, dass der C-Terminus des Peptids direkt vom Proteasom hergestellt wird und das auch TAP die C-terminale Aminosäure für den Transport benötigt. Bemerkenswert ist hier, dass die gleichen Aminosäuren an dieser Position von beiden Prozessen bevorzugt werden.

Die hier vorgestellte Art der stochastischen Modellierung eines Systems ist besser als die deterministische Modellierung des Systems mit gewöhnlichen Differentialgleichungen, da die hier betrachteten Prozesse eine große Variabilität aufweisen und die Teilchenzahlen gering sind. Durch das wiederholte simulieren des Systems erhält man eine gute Approximation des Zustandsraums, für den man sonst die aufwendigere Mastergleichung lösen müsste. Die Variabilität der Antigenpräsentation stellt eine interessante Herausforderung an die Präsentationskompetenz jeder einzelnen Zelle, denn sollte eine Zelle zu wenig oder keine Antigene präsentieren, hat ein infizierendes Agens die Möglichkeit, sich weiter im Organismus zu verbreiten und weiteren Schaden anzurichten. Die hier betrachteten Prozesse scheinen sicher zu stellen, dass immer genügend Peptide für die Antigenpräsentation zur Verfügung stehen.

Diese Ergebnisse wurden in Genome Informatics veröffentlicht[98].

5.2 Antigenprozessierung des HIV-1-BRU gag Proteins

In diesem Abschnitt soll das Modell auf ein physiologisch relevantes Substrat angewendet und der zeitliche Verlauf der Antigenpräsentation simuliert werden. Ziel der Simulation ist es die Infektion einer Zelle und die darauf folgende Antigenpräsentation darzustellen.

5.2.1 Charakterisierung des Proteins und des Modells

Das HIV-1-BRU gag Protein

Als Substrat wurde das gag Protein des HIV-1-BRU Virusstammes ausgewählt. Die Auswahl geschah zufällig mit der Forderung, dass es ein virales Protein sein soll, bei dem mindestens zwei 9mer Antigene[160, 175], die von dem Allel HLA-0201 präsentiert werden, in der SYFPEITHI Datenbank[106] enthalten sind.

Das Protein wird zur Assemblierung des Virus benötigt und enthält Strukturelemente[176].

Im folgenden Absatz ist die Sequenz des 512 Aminosäuren enthaltenden Proteins angegeben. In grün sind die beiden bekannten HLA-A0201 Epitope gekennzeichnet, in rot ist das in den folgenden Simulationen dominante 9mer Peptid hervorgehoben.

```
MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQL
QPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAADTGHSSQV
SQNYPIVQNIQGQMVHQAI SPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVG
GHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEIQIGWMTNNPPI
PVGEIYKRWII LGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQASQEVKNWMTET
LLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAAMSQVTNSATIMQRGNFR
NQRKIVKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNF
LQSRPEPTAPPFLQSRPEPTAPPEESFRSGVETTTPSQKQEPIDKELYPLTSLRSLFGNDPSSQ
```

Hierarchischer Aufbau des Modells

Das Modell wird in zwei Schritten entwickelt.

Im ersten Schritt wird das in Kapitel 4 entwickelte Modell des proteasomalen Abbaus von Proteinen und der damit verbundenen Fragmentbildung verwendet. Hierfür wird das Substrat 10'000'000 mal verdaut. Alle dabei entstehenden Fragmente erhalten so eine Rate beziehungsweise Wahrscheinlichkeit, mit der sie beim Abbau eines Proteins erzeugt werden.

Man muss dabei natürlich die Schwächen dieses Modells beachten und sich bewusst sein, dass in den Simulationen deutlich mehr Fragmente hergestellt werden, als experimentell beobachtet werden, jedoch existiert zur Zeit kein anderes Modell, das bessere Vorhersagen macht oder besser verwendet werden kann.

Der zweite Schritt enthält die Simulation des gesamten Pathways. Hierzu werden die Modelle der Proteinsynthese und Abbaus (Kapitel 2) inklusive der DRiP Bildung, die proteasomale Erzeugung von Peptiden (Kapitel 4) basierend auf den Schnittstellenvorhersagen für das Proteasom (Kapitel 3.3), der Transport der Peptide in das ER mittels TAP (Kapitel 3.2), das Trimmen und Abbauen von Peptiden im Cytosol und ER durch Aminopeptidasen (Kapitel 3.4) sowie die Bindung der Peptide an MHC Komplexe (Kapitel 3.1) miteinander kombiniert. Abbildung 43 gibt eine schematische Übersicht über alle simulierten Prozesse.

Wie in Kapitel 5.1.1 beschrieben wird eine stochastische Simulation durchgeführt, bei der die einzelnen Prozesse entsprechend ihrer Raten durchgeführt werden. Die Proteinsynthese und der Abbau werden nicht stochastisch simuliert.

Proteinsynthese und Abbau

Es werden die Gleichungen 1 und 2 verwendet, um den die Proteinsynthese und den proteasomalen Abbau zu beschreiben. Es gibt in diesem Fall keine Proteinreifung, somit ist der Parameter $k_3=0$. Es werden die Parameter k_{21} und k_{22} für NP-SIINFEKL-GFP verwendet. Für die mRNA Synthese (v_0) und den Abbau (b_0) sowie die Lebenszeit der DRiP Fraktion (a_1) werden die Parameter aus Tabelle 2 verwendet. Für die Lebenszeit des nativen Proteins wird eine Halbwertszeit von 24 Stunden ($a_2=4.81 \cdot 10^{-4} \text{ min}^{-1}$) angenommen. Mit den Gleichungen 1 und 2 kann zu jedem Zeitpunkt der Simulation berechnet werden, wie viele Proteine vom Proteasom abgebaut werden. Diese Werte werden anstatt der konstanten 20 beziehungsweise 200 Proteine pro Minute aus Kapitel 5.1.1 verwendet.

N-terminaler Abbau der Peptide im Cytosol und ER

Der schrittweise Abbau der Peptide in beiden Kompartimente wird wie im vorhergehenden Abschnitt gehandhabt. Die Raten der Aminopeptidasen werden entsprechend den in Kapitel 3.4 (Abbildung 28) bestimmten Werten gesetzt. Diese Werte werden allerdings noch so normiert, dass der Unterschied zwischen der Minimalrate und der Maximalrate der Faktor 100 ist und dass die relativen Unterschiede zwischen den Raten erhalten bleiben. Der durchschnittliche Abbau hat wieder eine Halbwertszeit von 5 Sekunden.

Bindung der Peptide an TAP und MHC

Die Bindung an TAP und MHC wird ebenfalls wie in Abschnitt 5.1.1 beschrieben durchgeführt. Die Peptide im Cytosol zwischen 8 und 19 Aminosäuren Länge werden mit dem TAP Modell (Kapitel 3.2) bewertet, das Durchschnittspeptid hat eine Halbwertszeit für den Transport von 120 Sekunden. Alle 9mer Peptide im ER werden mit dem in Kapitel 3.1

beschriebenen Modell von Peters et al. [111] bewertet, das Durchschnittspeptid hat eine Halbwertszeit von 40 Sekunden. Die Basisraten für TAP und MHC Bindung sind willkürlich festgelegt und müssen gegebenenfalls durch experimentelle Werte korrigiert werden.

Der Transport der MHC Komplexe zur Zelloberfläche wird nicht explizit modelliert und es gibt keinen Abbau der MHC Komplexe an der Zelloberfläche in den simulierten Zeiträumen.

5.2.2 Ergebnisse der Simulation

Es wurden mehrere 30 Minuten Simulationen mit dem oben beschriebenen Modell durchgeführt. In Abbildung 48 sind die zeitlichen Verläufe der beiden bekannten HLA-A0201 Epitope und des Peptids, das in den Simulationen mit der größten Menge präsentiert wird, dargestellt.

Das Peptid FLGKIWPSY wird in den Simulationen am häufigsten präsentiert. Es ist mit 17.3 Prozent das deutlich dominierend präsentierte Epitop. SLYNTVATL ist an Position 4 der präsentierten Epitope in der Simulation und kommt mit 5.5 Prozent relativ häufig vor. Das andere bekannte Epitop TLNAWVKVV ist an Position 31 und mit 0.6 Prozent eher selten aufzufinden.

Die Raten der Aminopeptidasen begünstigen die Bildung von FLGKIWPSY im Gegensatz zu gleichverteilten Raten, wie sie im vorhergehenden Abschnitt verwendet wurden mit einer Steigerung der Präsentation um 33.8 Prozent. SLYNTVATL wird mit einer Steigerung von 8.5 Prozent ebenfalls begünstigt, während bei TLNAWVKVV keine Änderung festzustellen ist.

Die Bindung von SLYNTVATL an den MHC Komplex ist mit 12 an MHC gebundenen Molekülen von 13 in das ER transportierten Precursoren deutlich effektiver als bei FLGKIWPSY (59 von 130), jedoch ist der Transport der FLGKIWPSY Precursoren in das ER mit zehnmal größerer Häufigkeit zu beobachten. Für TLNAWVKVV ist sowohl der Transport in das ER relativ niedrig als auch die Bindung an den MHC Komplex (2 von 7).

Auffallend ist das etwa die Hälfte aller FLGKIWPSY Precursoren als 10mer von TAP transportiert werden und das ausschließlich SLYNTVATL überhaupt als 9mer in das ER gelangt.

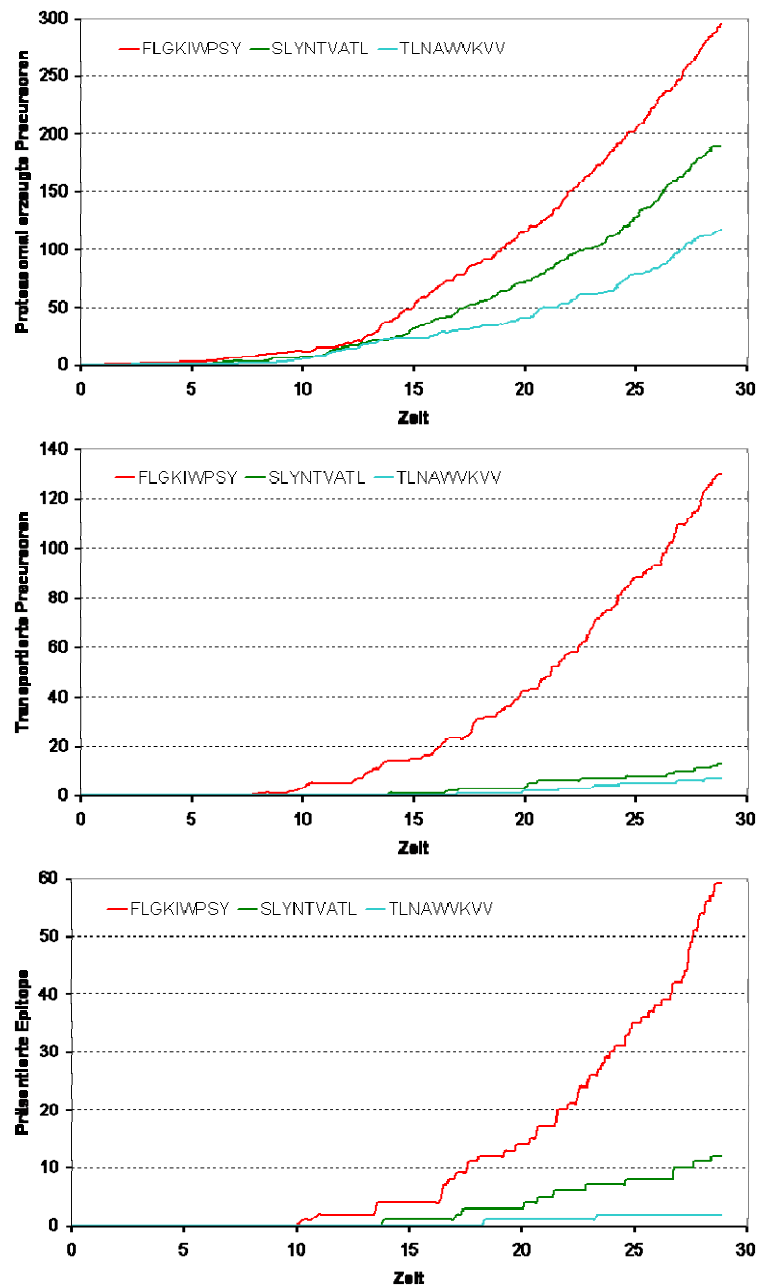


Abbildung 48: Simulationsverlauf ausgewählter Peptide aus HIV-1-BRU gag Protein

Die Abbildung zeigt die zeitlichen Verläufe der beiden bekannten Epitope und des in den Simulationen dominanten Peptids. Oben ist die proteasomale Produktion der Peptide und ihrer Precursoren bis zur Länge 20 dargestellt. Der angegebene Wert ist die Summe aller Precursoren. Die mittlere Abbildung zeigt, die Menge der in das ER transportierten Precursoren der drei Peptide. Unten ist die Anzahl der präsentierten 9mer Epitope dargestellt.

5.2.3 Diskussion

Vorhersagequalität des kinetischen Modells entspricht der der statistischen Modelle

An einem Beispiel kann natürlich keine statistisch verlässliche Aussage über Sensitivität und Spezifität des Modells getroffen werden. Wenn man aber die Werte berechnet erhält man zum einen eine Spezifität von 99.4 Prozent bei einer Sensitivität von 50 Prozent und eine Spezifität von 94.2 Prozent bei einer Sensitivität von 100 Prozent. Dieses ergibt einen AUC Wert von 0.968 und ist damit nicht schlechter, als die Vorhersagen, die beim Kombinieren der statistischen Modelle in Kapitel 3.5 gemacht werden können (Tabelle 19).

Detailliertes kinetisches Modell erlaubt das Verfolgen des Verlaufs jedes Peptids

Der große Vorteil des hier erstellten kinetischen Modells ist das Finden von Kandidaten für die Antigenpräsentation und die Möglichkeit jedes Peptid über die Simulation verfolgen zu können. So ist es möglich zu verstehen, durch welche Prozesse ein Peptid zum potentiellen Antigen wird und welche es vermindern. In dem hier gerechneten Beispiel ist es vor allem der TAP Transport, der das Peptid FLGKIWPSY so dominant gegenüber den anderen Epitopen macht wobei das aber durch eine hohe proteasomale Bildungsrate und die Aminopeptidasen noch unterstützt wird. Für SLYNTVATL ist vor allem die hohe Affinität zu den MHC Molekülen ausschlaggebend für die Präsenz an der Zelloberfläche. Durch diese Kenntnis der Details erhält man Hypothesen für weitere Experimente. So wäre es interessant zu verstehen, in welchem Schritt FLGKIWPSY nicht mehr prozessiert wird oder ob das Peptid einfach noch nicht als Antigen gefunden worden ist. So sind in der SYFPEITHI Datenbank neben MHC-Liganden auch T-Zell Epitope, also Peptide, die eine Antigenreaktion hervorrufen können enthalten. Die Erkennung durch T-Zellen ist aber kein Bestandteil des Modells und so ist es möglich das MHC Liganden nicht gefunden werden, weil sie keine Immunantwort auslösen können.

Die Alternative zur stochastischen Simulation des Zustandsraum, die Lösung der Mastergleichung, ist in diesem Beispiel schon nicht mehr möglich, da mehr als 500'000 Differentialgleichungen aufgestellt und numerisch integriert werden müssten.

Das Modell des MHC I Präsentationsweges gibt den Stand des Wissens wieder, hat aber auch noch Schwächen

Alle im Modell enthaltenen Teilmodelle sind auf dem Stand der Zeit beziehungsweise es gibt keine Modelle für den jeweiligen Abschnitt des MHC I Pathways, für den besser ge-

eignete Modelle existieren. Das Verhältnis der Basisraten der einzelnen Prozesse ist jedoch nicht experimentell bestimmt. So erscheint die Basisrate der MHC Bindung im Nachhinein als relativ hoch angesetzt, während die anderen Raten weiterhin physiologisch sinnvoll erscheinen. Es gibt leider keine experimentellen Messungen außer den angegebenen, um diese Raten festzulegen. Es bleibt aber festzuhalten, dass sich die relativen Verhältnisse der Peptide in einem Kompartiment sich nicht ändern, wenn die Basisraten verändert werden.

Für den Verlauf der Proteinmengen, die abgebaut werden, bräuchte man eigentlich die reale Lebenszeit des Proteins. Da aber die meisten viralen Proteine sehr stabil sind ist der Anteil der DRiP Fraktion dominierend innerhalb der Zeit direkt nach Infektion der Zellen (Abbildung 12) mit etwa 90% Antigenen aus DRiPs in diesem gerechneten Beispiel. Der Fakt, das HIV ein Retrovirus ist und nach der Infektion die virale RNA erst in die zelluläre DNA eingebaut werden muss, ist in diesem Modell nicht enthalten. Auch der Transport der fertigen MHC Komplexe aus dem ER über den Golgi Apparat an die Zelloberfläche ist nicht Bestandteil des Modells, wodurch sich Veränderungen auf der Zeitachse der Simulationen ergeben sollten.

Keiner der simulierten Prozesse wird als gesättigt modelliert beziehungsweise alle Peptide werden unabhängig voneinander betrachtet. Für TAP ist gezeigt worden, dass etwa ein Drittel aller Transporter durch Substrat besetzt sind[24]. In der Arbeit von Princiotta et al.[30] wurde gezeigt, dass die Anzahl der präsentierten Epitope durch die Menge des angebotenen Substrats und nicht durch die Verfügbarkeit der MHC Moleküle limitiert wird, da durch das Anbieten von MSIINFEKL die Antigenpräsentation um ein Vielfaches gesteigert wird. Die Aminopeptidasen im Cytosol und ER arbeiten sehr schnell und es ist schwer vorstellbar, dass sie im Sättigungsbereich arbeiten. In der in der Simulation betrachteten Zeit, findet auch kein nennenswerter Abbau der präsentierten MHC Komplexe statt. Die vereinfachenden Annahmen des Modells für diese Prozesse sind daher gerechtfertigt.

In der Arbeit von Princiotta et al.[30] wird abgeschätzt, dass fast alle Proteasome permanent mit dem Abbau von Proteinen beschäftigt sind. Hier ist also eine Konkurrenz der Substrate möglich, nur werden Aufnahme der Proteine und Wiederaufnahme von langen proteasomalen Produkten nicht modelliert. Insgesamt ist der gesamte Bereich der proteolytischen Peptiderstellung im Modell aufgrund fehlender experimenteller Informationen schlecht wiedergegeben. Zum einen werden deutlich mehr Peptide in der Simulation erzeugt, als experimentell gefunden werden (siehe Kapitel 4) und zum anderen fehlen andere Prozesse im Cytosol, wie die als Endo- und Exopeptidase arbeitende TPPII. Die Prozesse der Wiederaufnahme von Peptiden in das Proteasom und andere degradierende

Prozesse müssten mit Längenabhängigkeiten und Effektoren modelliert werden, was aber durch fehlende Daten zum Abbau von Proteinen *in vivo* zurzeit nicht möglich ist.

6 Zusammenfassung der Ergebnisse und Schlussfolgerungen

Das menschliche Immunsystem ist wichtig, um den Körper vor Pathogenen zu schützen. Zu diesem Zweck signalisieren die somatischen Zellen im Rahmen des MHC I Pathways den internen Proteinstatus indem sie kurze Peptide, die beim Abbau der Proteine entstehen, an der Zelloberfläche präsentieren.

Diese Arbeit soll durch die Entwicklung von Modellen und Durchführen von Simulationen das Verständnis des MHC I Pathways vertiefen und gegebenenfalls helfen die Vorhersage der präsentierten Antigene zu verbessern. In einem Teil dieser Arbeit soll die Frage erörtert werden, welchen Anteil die Präsentation der neu synthetisierten Proteine an der Antigenpräsentation hat.

Es wurden Modelle für die Teilbereiche des MHC I Pathways erstellt, sowie ein kinetisches Modell für die Synthese und den Abbau von viralen Proteinen. Diese Modelle wurden zu einem kinetischen Gesamtmodell kombiniert, mit dem die auf eine Infektion folgende Proteinsynthese und Antigenpräsentation simuliert werden kann.

Das Modell der Proteinsynthese und des Abbaus wurde mit experimentellen Daten spezieller Konstrukte kalibriert und die so erhaltenen Parameter ausgewertet. So ergibt sich, dass eine DRiP Rate von 10 Prozent ausreichend ist, um die beobachteten Zeitverläufe zu beschreiben. Diese Rate gewährleistet eine Antigenpräsentation unabhängig von der Lebenszeit des Proteins und beschleunigt Immunpräsentation vor allem bei langlebigen Proteinen, was bei viralen Proteinen oftmals der Fall ist. Für langlebige Proteine gilt daher, dass vor allem die Proteinsynthese und nur in geringem Maß der Abbau der nativen Proteine ausschlaggebend für die Antigenpräsentation ist. Bei kurzlebigen Proteinen ist der Anteil der Antigene aus DRiPs deutlich geringer. Weiterhin zeigt sich, dass der Abbau mit Kinetiken erster Ordnung ausreichend gut beschrieben werden kann, solange keine gesonderten Mechanismen für die gezielte Steuerung des Abbaus des speziellen Proteins existieren.

Für die Bindung der zu präsentierenden Moleküle an MHC Moleküle sind gute Modelle etabliert, die zeigen, dass dieser Schritt der am stärksten selektive Teil des gesamten Pathways ist. Eins dieser existierenden Modelle wurde für die Pathwaysimulation und die Kombination von Teilmodellen verwendet.

Es wurde ein Teilmodell für die proteasomale Schnittvorhersage erstellt. Grundannahme ist, dass alle Aminosäuren unabhängig voneinander zur Hydrolyserate der einzelnen Bindungen beitragen. Das zehn Aminosäuren umfassende Fenster von P6 bis P4' trägt signi-

fikant zur Vorhersage von proteasomalen Schnitten bei, wobei die Beiträge von P1 und P1' dominant sind. Die Vorhersagequalität von C-terminalen Schnitten von Epitopen ist insgesamt gering mit einem AUC Wert von 0.67 beim konstitutiven und 0.76 beim Immunoproteasom. Somit hat das Proteasom entweder eine geringe Präferenz für spezielle Motive und damit geringe Spezifität oder es determinieren Effekte die Schnittrate, die sich nicht mit der Annahme der unabhängigen Beiträge der Aminosäuren im Sequenzfenster vereinbaren lassen und auch nicht mit so einem Modell abzubilden sind. Der proteasomale Abbau von Proteinen ist experimentell schwer zugänglich. So müssen die gebildeten Fragmente aufwändig identifiziert und quantifiziert werden, während bei den anderen Schritten des Antigenpräsentationswegs Bindungsstudien ausreichend sind. Daher existieren wenige Daten zu diesem Teil des MHC I Pathways und es gibt keine Möglichkeit Paarkoeffizienten oder andere Modelle wie ein zweischichtiges neuronales Netzwerk ohne Overfitting zu trainieren. Es gibt starke Hinweise darauf, dass nicht alle C-terminalen Schnitte der präsentierten Epitope vom Proteasom hergestellt werden. Vor allem Peptide, die ein Lysin an der P1 Position haben (P1K Liganden), werden mit hoher Wahrscheinlichkeit von der TPPII geschnitten. Das immuno und konstitutive Proteasom stellen unterschiedliche Fragmente her und auch die Vorhersagequalität der beiden Proteasomtypen ist unterschiedlich. Für die Antigenpräsentation ist der Effekt aber fast identisch, da die vom konstitutiven Proteasom hergestellten zusätzlichen Fragmente von TAP nur mit geringer Effizienz transportiert werden, wodurch die Vorhersagequalität für beide Proteasommodelle in Kombination mit TAP bei 0.83 ist. Dieses zeigt, dass TAP beim Transport der Peptide eine ähnliche Präferenz, wie das Immunoproteasom beim Schneiden hat (Korrelation $r^2=0.37$). Somit stellt das Immunoproteasom mehr Peptide her, die für die Antigenpräsentation geeignet sind.

Es wurde ein Teilmodell für den Transport von Peptiden in das ER mittels TAP erstellt. Es konnte damit gezeigt werden, dass Precursoren von Epitopen wichtig sind und das Einbeziehen eine Verbesserung der Epitopvorhersage bewirkt. Das erstellte Modell beruht auf der Annahme der unabhängigen Beiträge der einzelnen Aminosäuren und die Qualität der Modellvorhersagen sowie die nur geringe Verbesserung durch Einschließen von Paarkoeffizienten rechtfertigt diese Annahme. Das Modell ist in der Lage Transportraten für Peptide unterschiedlicher Längen vorherzusagen. In Kombination mit dem Proteasommodell können so Epitope für alle humanen MHC I Allele mit guter Qualität vorhergesagt werden. Bei der Vorhersage von Epitopen für verschiedene MHC Allele zeigt sich, dass einzelne Allele den von TAP und Proteasom präferenzierten C-Terminus ebenfalls bevorzugen (z.B. HLA B-35, HLA B-07, HLA B-08), während andere C-terminal eine andere Präferenz haben und besser mit dem von TAP transportierten N-Terminus vorhergesagt werden können (HLA B-27, HLA A-03). Hierbei bleibt anzumerken, dass

HLA A-03 gut P1K Liganden bindet, die vermutlich nicht vom Proteasom hergestellt werden.

Das Teilmodell der cytosolischen und endoplasmatischen Aminopeptidasen besteht im Wesentlichen darin, dass von jedem Peptid im entsprechenden Kompartiment die N-terminale Aminosäure entfernt werden kann. Die entsprechenden Raten sind von experimentellen Partnern an Modellsubstraten gemessen worden. Die Analyse der N-terminal flankierenden Region von MHC I Epitopen zeigt, dass lange Precursoren weder von cytosolischen noch von endoplasmatischen Aminopeptidasen gekürzt werden, also beispielsweise TPPII als Endopeptidase arbeitet. Precursoren mit einer Länge zwischen 13 und 17 Aminosäuren werden präferenziell im Cytosol getrimmt. Erst kurze Precursoren mit einer Länge bis 12 Aminosäuren haben eine höhere Wahrscheinlichkeit im ER gekürzt zu werden. Dieses zeigt, dass die Aminopeptidasen einen Einfluss auf die Antigenpräsentation da der N-terminale Schutz der Epitope durch die N-terminale Verlängerung im Cytosol nicht zerstört wird und erst im ER das Epitop auf seine finale Länge getrimmt wird.

Alle hier erstellten Teilmodelle des MHC I Pathways wurden auf *in vitro* Daten spezifisch für den jeweiligen Schritt trainiert. Nur so ist es möglich, die einzelnen Schritte zu verstehen und die relativen Beiträge der einzelnen Abschnitte zu wichten. Jedes Modell kann als isoliertes Modul verwendet werden, um Vorhersagen zu machen oder mit anderen Methoden kombiniert zu werden. Modelle, die mit *in vivo* Daten trainiert worden sind, sind gegebenenfalls in der Lage gute Vorhersagen zu liefern, tragen aber nicht zum Verständnis der einzelnen Schritte bei. Die Interaktion der einzelnen Schritte des Pathways kann natürlich erst verstanden werden, wenn die Module miteinander kombiniert werden.

Die Kombination der Module im MHC I Pathway gibt sehr hohe AUC Werte abhängig von der Qualität der MHC I Molekül Bindungsvorhersage zwischen 0.951 und 0.997. Somit hat die Kombination von *in vitro* Modellen von TAP Transport, proteasomalem Schneiden des C-Terminus und MHC I Bindung die gleiche oder eine höhere Vorhersagequalität als *in vivo* basierte Modelle wie BIMAS[160] oder SYFPEITHY[106] und tragen weiterhin zum Verständnis der zugrunde liegenden Prozesse bei. Es ist nicht trivial, das richtige Protokoll für die Kombination der statistischen Modelle zu finden, da die Modelle zueinander gewichtet werden müssen und bei den erreichten hohen Vorhersagequalitäten falsch negativ Vorhersagen in einem Teilmodell die Gesamtqualität deutlich senken. Diese Wichtung der relativen Beiträge der Module muss auf die Fragestellung angepasst werden, je nachdem, ob man einige der realen Epitope finden möchte mit möglichst wenigen falsch positiven Vorhersagen oder ob man alle Epitope finden möchte.

Die statistischen Modelle für proteasomale Schnittpräferenz, TAP Transport und MHC I Molekül Bindung für einige Allele wurden in einem Online Tool verfügbar gemacht

(<http://mhc-pathway.net>). Diese Vorhersagetools sind mittlerweile zur IEDB Internetseite migriert worden, wo sie mit weiteren MHC I Molekül Bindungsvorhersagen kombiniert werden und aktuell gehalten werden (<http://tools.immuneepitope.org>). Die Vorhersage ist auf http://tools.immuneepitope.org/analyze/html/mhc_processing.html zu finden.

Auf Basis der proteasomalen Schnittstellenvorhersage wurde ein einfaches Modell der proteasomalen Fragmenterstellung erzeugt. Das „slide & slice Modell“ basiert auf den Annahmen, dass das Substrat sukzessiv durch das Proteasom transportiert wird und an jeder Peptidbindung entsprechend der Schnittwahrscheinlichkeit und der Länge des erzeugten Fragments entschieden wird, ob das Protein um eine Aminosäure weiter in das Proteasom wandert oder geschnitten wird. Die Aufnahme des Substrats in das Proteasom wird nicht modelliert und die Parameter so eingestellt, dass ein weiteres Hydrolysieren bereits erzeugter Fragmente nicht mehr notwendig ist. Das Modell wurde so parametrisiert, dass die experimentell beobachtete Längenverteilung von Fragmenten sich gut reproduzieren lässt. Die maximale Schnittrate des Proteasoms liegt im Bereich der minimalen Bewegungsrate des Substrates. Damit hat selbst die Aminosäuresequenz mit höchstmöglicher Schnittwahrscheinlichkeit eine Rate von weniger als 50 Prozent, um geschnitten zu werden. Die exakte Längenverteilung ist im Experiment von der Aufnahme rate des Proteins und der entstandenen Fragmente sowie der Dauer des Experiments abhängig. Die Fragmentvorhersage ist etwas besser als das zufällige Ziehen von Peptiden aus der richtigen Längenverteilung. Allerdings werden etwa 40 bei einer Sensitivität von 50 Prozent bis 100 bei einer Sensitivität von 100 Prozent mal mehr Peptide vorhergesagt, als gefunden werden. Der AUC Wert der Vorhersage ist mit 0.92 ist sehr hoch, was aber an der Verwendung einer geeigneten Längenverteilung liegt, was ein sehr einfaches Klassifizieren einer Vielzahl der möglichen Proteinfragmente ermöglicht. Die geringe Qualität der Fragmentvorhersage konnte durch das Einbeziehen von Sekundärstrukturen nicht verbessert werden und diese Strukturen haben keinen Einfluss auf die Schnittraten im Proteasom. Andere Effekte entweder aus Kombinationen lokaler Sequenzmotive oder allosterische Effekte müssen offenbar definieren, ob ein Schnitt stattfindet. Es werden jedoch experimentell sicher nicht alle hergestellten Fragmente gefunden. Es gibt zurzeit kein zufrieden stellendes Modell der proteasomalen Fragmenterstellung.

Im letzten Abschnitt dieser Arbeit wurden die erstellten Modelle zur Antigenpräsentation in einem mechanistischen kinetischen Modell miteinander kombiniert, um die Infektion einer Zelle durch einen Virus zu simulieren. Es wird mit einem stochastischen Modell Trajektorien im Zustandsraum simuliert. Durch ausreichend viele dieser Trajektorien kann die Lösung der Mastergleichung approximiert werden. Es wurde zum einen ein Peptid aus dem Protein pp89, für das die proteasomale Fragmenterstellung experimentell gemessen wurde, verwendet. Das Modell bestand aus der Kombination cytosolischer Amino-peptida-

sen mit proteasomaler Fragmentproduktion und TAP vermitteltem Transport der Peptide in das ER. Es wurde gezeigt, dass Aminopeptidasen das Spektrum der Substrate für TAP Erhöhen indem sie weitere Precursoren herstellen, als vom Proteasom erstellt werden. Precursoren bilden zusätzlich natürlich einen Schutz für das Epitop vor cytosolischem Abbau.

In einem zweiten Beispiel wurde die Antigenpräsentation aus dem gag Protein des HIV-1-BRU Virus simuliert. Es werden die Modelle für Proteinsynthese und Abbau, proteasomaler Fragmenterstellung, cytosolischen und endoplasmatischen N-terminalen Peptidasen, TAP vermitteltem Transport und MHC-Bindung miteinander kombiniert. Für diese Art der Kombination ist die Kenntnis der relativen Basisraten wichtig, um die Modellzeit realistisch zu halten, aber die prädiktive Qualität wird davon nicht beeinflusst. Die Vorhersagequalität liegt im Bereich der kombinierten statistischen Modelle mit dem Vorteil von quantitativen Aussagen, stellt jedoch hohe Anforderungen an die Rechenkapazität. Es kann im Prinzip jedes Peptid über den Verlauf der Simulation verfolgt werden und verstanden werden, in welchen Schritten des Pathways das Peptid zu einem potentiellen Epitop wird oder in welchem Schritt es von der Antigenpräsentation ausgeschlossen wird. Es zeigt sich, dass obwohl die MHC Bindung der am stärksten selektive Schritt im MHC I Pathway ist, aber außerdem jeder der betrachteten Teilprozesse einen verstärkenden oder vermindernenden Einfluss auf die Präsentation eines Epitops haben kann. So wird das am stärksten präsentierte Peptid des Modellsubstrats zum dominanten Epitop aufgrund der günstigen Raten von Proteasom, TAP und der Aminopeptidasen und nicht aufgrund der Affinität zum MHC I Molekül. Die so erhaltenen Simulationsergebnisse können experimentell validiert werden und dadurch beim Aufdecken von Widersprüchen das Verständnis der Funktionsweise des Pathways vertiefen.

Im Review von Kloetzel[32] wird die Ansicht vertreten, dass das Proteasom primär sehr lange Fragmente erstellt, die dann zunächst in das Cytosol freigesetzt werden, wo sie von TPPII und anderen Peptidasen bearbeitet werden und dann nochmals das Proteasom passieren, um auf ihre endgültige Länge geschnitten zu werden. Um diese Prozesse mit einem Modell realistisch abbilden zu können sind bisher nicht vorhandene experimentelle Daten notwendig. So benötigt man die Raten bzw. Wahrscheinlichkeiten, mit der Peptide oder Fragmente vom Proteasom aufgenommen werden und auch wieder freigesetzt werden. Auch die anderen beitragenden Peptidasen müssten besser charakterisiert werden in Bezug auf ihre Substratspezifität.

Zusammenfassend bleibt festzuhalten, dass die Vorhersage von Antigenen mit hoher Qualität funktioniert. Die aufgestellten Teilmodelle und deren Kombination zeigen, dass keine weiteren Schritte, als die bereits bekannten notwendig sind, um die Antigenpräsen-

tation zu beschreiben. Durch Aufstellen des Modells und Simulation von viralen Infektionen zeigt sich, dass die größte Unsicherheit beziehungsweise Wissensdefizit im Bereich der proteasomalen Fragmenterstellung und dem Schicksal längerer Peptide im Cytosol ist. Durch neue Experimente können die erstellten Modelle validiert und weiter verbessert werden, was letztendlich zu einem besseren Verständnis des Pathways, der Interaktion der beteiligten Prozesse und damit zu einer Verbesserung der Vorhersagequalität führt.

7 Ausblick

Um dem endgültigen Ziel, einem zeitlich aufgelösten Modell der Antigenpräsentation, entscheidend näher zu kommen, muss die Frage der Peptidgenerierung mittels Proteasom besser geklärt werden. Obwohl die Vorhersage von Schnittstellen viel versprechend funktioniert, wurde gezeigt, dass die Vorhersage von Proteinfragmenten nur wenig besser als zufällig arbeitet. Dies deutet darauf hin, dass die Information, die für die Erzeugung von Peptiden verantwortlich ist, nicht allein in der Sequenz des Proteins nahe der Schnittstelle liegt. Es bedarf meiner Meinung nach eines mechanistischen Bildes der Translokation des Proteins durch das Proteasom, um ein verlässliches Modell für die Fragmentbildung zu erzeugen. Hierzu könnte man die Struktur des Proteasoms mit einem Proteinsubstrat an mehreren Stellen des Prozesses analysieren. Für ein eher statistisches statt mechanistisches Modell würde man zahlreiche Proteinverdaus benötigen, um aus der gesamten Sequenz der Proteine und möglicherweise auch ihrer Struktur Informationen für die Generierung von Fragmenten zu extrahieren.

Ein weiterer Punkt die proteasomale Abbaurate der Proteine, da diese natürlich die verfügbare Menge an Substrat für die Antigenpräsentation bestimmt. Zurzeit ist nur für eine geringe Menge von Proteinen die tatsächliche Lebenszeit sowie der Mechanismus der Steuerung des Abbaus (Ubiquitin System gesteuert, Abbau nach mehreren Schädigungen oder Abbau erster Ordnung) bekannt. Hier konnte jedoch gezeigt werden, dass durch den Abbau eines kleinen Teils der synthetisierten Proteine (DRiPs) die Zelle in der Lage ist Antigene aller synthetisierten Proteine weitgehend unabhängig von ihren Lebenszeiten zu präsentieren und dabei die Menge der präsentierten Epitope proportional zur Menge des hergestellten Proteins ist.

Literatur

- [1] Pamer, E. und Cresswell, P. (1998): Mechanisms of MHC class I--restricted antigen processing, *Annu Rev Immunol* 16, Seite 323-58.
- [2] Rock, K. L. und Goldberg, A. L. (1999): Degradation of cell proteins and the generation of MHC class I-presented peptides, *Annu Rev Immunol* 17, Seite 739-79.
- [3] Shastri, N.; Schwab, S. und Serwold, T. (2002): Producing nature's gene-chips: the generation of peptides for display by MHC class I molecules, *Annu Rev Immunol* 20, Seite 463-93.
- [4] York, I. A. und Rock, K. L. (1996): Antigen processing and presentation by the class I major histocompatibility complex, *Annu Rev Immunol* 14, Seite 369-96.
- [5] (1999): Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium, *Nature* 401 [6756], Seite 921-3.
- [6] Vandiedonck, C. und Knight, J. C. (2009): The human Major Histocompatibility Complex as a paradigm in genomics research, *Brief Funct Genomic Proteomic* 8 [5], Seite 379-94.
- [7] Marsh, S. G.; Albert, E. D.; Bodmer, W. F.; Bontrop, R. E.; Dupont, B.; Erlich, H. A.; Geraghty, D. E.; Hansen, J. A.; Hurley, C. K.; Mach, B.; Mayr, W. R.; Parham, P.; Petersdorf, E. W.; Sasazuki, T.; Schreuder, G. M.; Strominger, J. L.; Svejgaard, A.; Terasaki, P. I. und Trowsdale, J. (2005): Nomenclature for factors of the HLA system, 2004, *Tissue Antigens* 65 [4], Seite 301-69.
- [8] WHO (2010): Offizielle HLA-Nomenklatur der WHO, 2010
- [9] Jinek, M. und Doudna, J. A. (2009): A three-dimensional view of the molecular machinery of RNA interference, *Nature* 457 [7228], Seite 405-12.
- [10] Siomi, H. und Siomi, M. C. (2009): On the road to reading the RNA-interference code, *Nature* 457 [7228], Seite 396-404.
- [11] Houseley, J.; LaCava, J. und Tollervey, D. (2006): RNA-quality control by the exosome, *Nat Rev Mol Cell Biol* 7 [7], Seite 529-39.
- [12] Vanacova, S. und Stefl, R. (2007): The exosome and RNA quality control in the nucleus, *EMBO Rep* 8 [7], Seite 651-7.
- [13] Kapp, L. D. und Lorsch, J. R. (2004): The molecular mechanics of eukaryotic translation, *Annu Rev Biochem* 73, Seite 657-704.
- [14] Deuerling, E. und Bukau, B. (2004): Chaperone-assisted folding of newly synthesized proteins in the cytosol, *Crit Rev Biochem Mol Biol* 39 [5-6], Seite 261-77.
- [15] Young, J. C.; Agashe, V. R.; Siegers, K. und Hartl, F. U. (2004): Pathways of chaperone-mediated protein folding in the cytosol, *Nat Rev Mol Cell Biol* 5 [10], Seite 781-91.
- [16] Hershko, A. und Ciechanover, A. (1998): The ubiquitin system, *Annu Rev Biochem* 67, Seite 425-79.
- [17] Liu, Y. C. (2004): Ubiquitin ligases and the immune response, *Annu Rev Immunol* 22, Seite 81-127.
- [18] Bachmair, A.; Finley, D. und Varshavsky, A. (1986): In vivo half-life of a protein is a function of its amino-terminal residue, *Science* 234 [4773], Seite 179-86.
- [19] Belle, A.; Tanay, A.; Bitincka, L.; Shamir, R. und O'Shea, E. K. (2006): Quantification of protein half-lives in the budding yeast proteome, *Proc Natl Acad Sci U S A* 103 [35], Seite 13004-9.
- [20] Doherty, M. K.; Hammond, D. E.; Clague, M. J.; Gaskell, S. J. und Beynon, R. J. (2009): Turnover of the human proteome: determination of protein intracellular stability by dynamic SILAC, *J Proteome Res* 8 [1], Seite 104-12.

- [21] Yewdell, J. W.; Anton, L. C. und Bennink, J. R. (1996): Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules?, *J Immunol* 157 [5], Seite 1823-6.
- [22] Schubert, U.; Anton, L. C.; Gibbs, J.; Norbury, C. C.; Yewdell, J. W. und Bennink, J. R. (2000): Rapid degradation of a large fraction of newly synthesized proteins by proteasomes, *Nature* 404 [6779], Seite 770-4.
- [23] Khan, S.; de Giuli, R.; Schmidtke, G.; Bruns, M.; Buchmeier, M.; van den Broek, M. und Groettrup, M. (2001): Cutting edge: neosynthesis is required for the presentation of a T cell epitope from a long-lived viral protein, *J Immunol* 167 [9], Seite 4801-4.
- [24] Reits, E. A.; Vos, J. C.; Gromme, M. und Neefjes, J. (2000): The major substrates for TAP in vivo are derived from newly synthesized proteins, *Nature* 404 [6779], Seite 774-8.
- [25] Turner, G. C. und Varshavsky, A. (2000): Detecting and measuring cotranslational protein degradation in vivo, *Science* 289 [5487], Seite 2117-20.
- [26] Schwab, S. R.; Li, K. C.; Kang, C. und Shastri, N. (2003): Constitutive display of cryptic translation products by MHC class I molecules, *Science* 301 [5638], Seite 1367-71.
- [27] Markossian, K. A. und Kurganov, B. I. (2004): Protein folding, misfolding, and aggregation. Formation of inclusion bodies and aggresomes, *Biochemistry (Mosc)* 69 [9], Seite 971-84.
- [28] Borges, J. C. und Ramos, C. H. (2005): Protein folding assisted by chaperones, *Protein Pept Lett* 12 [3], Seite 257-61.
- [29] Yewdell, J. W. und Nicchitta, C. V. (2006): The DRiP hypothesis decennial: support, controversy, refinement and extension, *Trends Immunol* 27 [8], Seite 368-73.
- [30] Princiotta, M. F.; Finzi, D.; Qian, S. B.; Gibbs, J.; Schuchmann, S.; Buttgerit, F.; Bennink, J. R. und Yewdell, J. W. (2003): Quantitating protein synthesis, degradation, and endogenous antigen processing, *Immunity* 18 [3], Seite 343-54.
- [31] Yewdell, J. W.; Reits, E. und Neefjes, J. (2003): Making sense of mass destruction: quantitating MHC class I antigen presentation, *Nat Rev Immunol* 3 [12], Seite 952-61.
- [32] Kloetzel, P. M. (2004): Generation of major histocompatibility complex class I antigens: functional interplay between proteasomes and TPPII, *Nat Immunol* 5 [7], Seite 661-9.
- [33] Groll, M.; Ditzel, L.; Lowe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D. und Huber, R. (1997): Structure of 20S proteasome from yeast at 2.4 Å resolution, *Nature* 386 [6624], Seite 463-71.
- [34] Elliott, T. (1997): Transporter associated with antigen processing, *Adv Immunol* 65, Seite 47-109.
- [35] Lankat-Buttgereit, B. und Tampe, R. (1999): The transporter associated with antigen processing TAP: structure and function, *FEBS Lett* 464 [3], Seite 108-12.
- [36] Lankat-Buttgereit, B. und Tampe, R. (2002): The transporter associated with antigen processing: function and implications in human diseases, *Physiol Rev* 82 [1], Seite 187-204.
- [37] Jones, P. M. und George, A. M. (1999): Subunit interactions in ABC transporters: towards a functional architecture, *FEMS Microbiol Lett* 179 [2], Seite 187-202.
- [38] Schneider, E. und Hunke, S. (1998): ATP-binding-cassette (ABC) transport systems: functional and structural aspects of the ATP-hydrolyzing subunits/domains, *FEMS Microbiol Rev* 22 [1], Seite 1-20.

- [39] Gaudet, R. und Wiley, D. C. (2001): Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing, *Embo J* 20 [17], Seite 4964-72.
- [40] Velarde, G.; Ford, R. C.; Rosenberg, M. F. und Powis, S. J. (2001): Three-dimensional structure of transporter associated with antigen processing (TAP) obtained by single Particle image analysis, *J Biol Chem* 276 [49], Seite 46054-63.
- [41] Armandola, E. A.; Momburg, F.; Nijenhuis, M.; Bulbuc, N.; Fruh, K. und Hammerling, G. J. (1996): A point mutation in the human transporter associated with antigen processing (TAP2) alters the peptide transport specificity, *Eur J Immunol* 26 [8], Seite 1748-55.
- [42] Momburg, F.; Armandola, E. A.; Post, M. und Hammerling, G. J. (1996): Residues in TAP2 peptide transporters controlling substrate specificity, *J Immunol* 156 [5], Seite 1756-63.
- [43] Nijenhuis, M.; Schmitt, S.; Armandola, E. A.; Obst, R.; Brunner, J. und Hammerling, G. J. (1996): Identification of a contact region for peptide on the TAP1 chain of the transporter associated with antigen processing, *J Immunol* 156 [6], Seite 2186-95.
- [44] Neumann, L. und Tampe, R. (1999): Kinetic analysis of peptide binding to the TAP transport complex: evidence for structural rearrangements induced by substrate binding, *J Mol Biol* 294 [5], Seite 1203-13.
- [45] Androlewicz, M. J.; Ortmann, B.; van Endert, P. M.; Spies, T. und Cresswell, P. (1994): Characteristics of peptide and major histocompatibility complex class I/beta 2-microglobulin binding to the transporters associated with antigen processing (TAP1 and TAP2), *Proc Natl Acad Sci U S A* 91 [26], Seite 12716-20.
- [46] van Endert, P. M.; Lopez, M. T.; Patel, S. D.; Monaco, J. J. und McDevitt, H. O. (1992): Genomic polymorphism, recombination, and linkage disequilibrium in human major histocompatibility complex-encoded antigen-processing genes, *Proc Natl Acad Sci U S A* 89 [23], Seite 11594-7.
- [47] Ma, W.; Lehner, P. J.; Cresswell, P.; Pober, J. S. und Johnson, D. R. (1997): Interferon-gamma rapidly increases peptide transporter (TAP) subunit expression and peptide transport capacity in endothelial cells, *J Biol Chem* 272 [26], Seite 16585-90.
- [48] Daniel, S.; Caillat-Zucman, S.; Hammer, J.; Bach, J. F. und van Endert, P. M. (1997): Absence of functional relevance of human transporter associated with antigen processing polymorphism for peptide selection, *J Immunol* 159 [5], Seite 2350-7.
- [49] Quadri, S. A. und Singal, D. P. (1998): Peptide transport in human lymphoblastoid and tumor cells: effect of transporter associated with antigen presentation (TAP) polymorphism, *Immunol Lett* 61 [1], Seite 25-31.
- [50] Momburg, F. und Tan, P. (2002): Tapasin-the keystone of the loading complex optimizing peptide binding by MHC class I molecules in the endoplasmic reticulum, *Mol Immunol* 39 [3-4], Seite 217-33.
- [51] Ahn, K.; Meyer, T. H.; Uebel, S.; Sempe, P.; Djaballah, H.; Yang, Y.; Peterson, P. A.; Fruh, K. und Tampe, R. (1996): Molecular mechanism and species specificity of TAP inhibition by herpes simplex virus ICP47, *Embo J* 15 [13], Seite 3247-55.
- [52] Tomazin, R.; Hill, A. B.; Jugovic, P.; York, I.; van Endert, P.; Ploegh, H. L.; Andrews, D. W. und Johnson, D. C. (1996): Stable binding of the herpes simplex virus ICP47 protein to the peptide binding site of TAP, *Embo J* 15 [13], Seite 3256-66.
- [53] Li, Y.; Salter-Cid, L.; Vitiello, A.; Preckel, T.; Lee, J. D.; Angulo, A.; Cai, Z.; Peterson, P. A. und Yang, Y. (2000): Regulation of transporter associated with antigen processing by phosphorylation, *J Biol Chem* 275 [31], Seite 24130-5.

- [54] Ortmann, B.; Copeman, J.; Lehner, P. J.; Sadasivan, B.; Herberg, J. A.; Grandea, A. G.; Riddell, S. R.; Tampe, R.; Spies, T.; Trowsdale, J. und Cresswell, P. (1997): A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes, *Science* 277 [5330], Seite 1306-9.
- [55] Celie, P. H.; Toebes, M.; Rodenko, B.; Ovaa, H.; Perrakis, A. und Schumacher, T. N. (2009): UV-induced ligand exchange in MHC class I protein crystals, *J Am Chem Soc* 131 [34], Seite 12298-304.
- [56] Garboczi, D. N.; Ghosh, P.; Utz, U.; Fan, Q. R.; Biddison, W. E. und Wiley, D. C. (1996): Structure of the complex between human T-cell receptor, viral peptide and HLA-A2, *Nature* 384 [6605], Seite 134-41.
- [57] Beninga, J.; Rock, K. L. und Goldberg, A. L. (1998): Interferon-gamma can stimulate post-proteasomal trimming of the N terminus of an antigenic peptide by inducing leucine aminopeptidase, *J Biol Chem* 273 [30], Seite 18734-42.
- [58] Kloetzel, P. M. und Ossendorp, F. (2004): Proteasome and peptidase function in MHC-class-I-mediated antigen presentation, *Curr Opin Immunol* 16 [1], Seite 76-81.
- [59] Levy, F.; Burri, L.; Morel, S.; Peitrequin, A. L.; Levy, N.; Bachi, A.; Hellman, U.; Van den Eynde, B. J. und Servis, C. (2002): The final N-terminal trimming of a subaminoterminal proline-containing HLA class I-restricted antigenic peptide in the cytosol is mediated by two peptidases, *J Immunol* 169 [8], Seite 4161-71.
- [60] Saric, T.; Beninga, J.; Graef, C. I.; Akopian, T. N.; Rock, K. L. und Goldberg, A. L. (2001): Major histocompatibility complex class I-presented antigenic peptides are degraded in cytosolic extracts primarily by thimet oligopeptidase, *J Biol Chem* 276 [39], Seite 36474-81.
- [61] Stoltze, L.; Schirle, M.; Schwarz, G.; Schroter, C.; Thompson, M. W.; Hersh, L. B.; Kalbacher, H.; Stevanovic, S.; Rammensee, H. G. und Schild, H. (2000): Two new proteases in the MHC class I processing pathway, *Nat Immunol* 1 [5], Seite 413-8.
- [62] York, I. A.; Mo, A. X.; Lemerise, K.; Zeng, W.; Shen, Y.; Abraham, C. R.; Saric, T.; Goldberg, A. L. und Rock, K. L. (2003): The cytosolic endopeptidase, thimet oligopeptidase, destroys antigenic peptides and limits the extent of MHC class I antigen presentation, *Immunity* 18 [3], Seite 429-40.
- [63] Saric, T.; Chang, S. C.; Hattori, A.; York, I. A.; Markant, S.; Rock, K. L.; Tsujimoto, M. und Goldberg, A. L. (2002): An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides, *Nat Immunol* 3 [12], Seite 1169-76.
- [64] Serwold, T.; Gonzalez, F.; Kim, J.; Jacob, R. und Shastri, N. (2002): ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum, *Nature* 419 [6906], Seite 480-3.
- [65] York, I. A.; Chang, S. C.; Saric, T.; Keys, J. A.; Favreau, J. M.; Goldberg, A. L. und Rock, K. L. (2002): The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues, *Nat Immunol* 3 [12], Seite 1177-84.
- [66] Day, P. M.; Yewdell, J. W.; Porgador, A.; Germain, R. N. und Bennink, J. R. (1997): Direct delivery of exogenous MHC class I molecule-binding oligopeptides to the endoplasmic reticulum of viable cells, *Proc Natl Acad Sci U S A* 94 [15], Seite 8064-9.
- [67] Gromme, M. und Neefjes, J. (2002): Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways, *Mol Immunol* 39 [3-4], Seite 181-202.
- [68] Marusina, K.; Reid, G.; Gabathuler, R.; Jefferies, W. und Monaco, J. J. (1997): Novel peptide-binding proteins and peptide transport in normal and TAP-deficient microsomes, *Biochemistry* 36 [4], Seite 856-63.

- [69] Yewdell, J. W. und Princiotta, M. F. (2004): Proteasomes get by with lots of help from their friends, *Immunity* 20 [4], Seite 362-3.
- [70] Begley, G. S.; Horvath, A. R.; Taylor, J. C. und Higgins, C. F. (2005): Cytoplasmic domains of the transporter associated with antigen processing and P-glycoprotein interact with subunits of the proteasome, *Mol Immunol* 42 [1], Seite 137-41.
- [71] Anton, L. C.; Schubert, U.; Bacik, I.; Princiotta, M. F.; Wearsch, P. A.; Gibbs, J.; Day, P. M.; Realini, C.; Rechsteiner, M. C.; Bennink, J. R. und Yewdell, J. W. (1999): Intracellular localization of proteasomal degradation of a viral antigen, *J Cell Biol* 146 [1], Seite 113-24.
- [72] Jackson, S. E.; Craggs, T. D. und Huang, J. R. (2006): Understanding the folding of GFP using biophysical techniques, *Expert Rev Proteomics* 3 [5], Seite 545-59.
- [73] Pakhomov, A. A. und Martynov, V. I. (2008): GFP family: structural insights into spectral tuning, *Chem Biol* 15 [8], Seite 755-64.
- [74] Heim, R.; Cubitt, A. B. und Tsien, R. Y. (1995): Improved green fluorescence, *Nature* 373 [6516], Seite 663-4.
- [75] Kloetzel, P. M. (2004): The proteasome and MHC class I antigen processing, *Biochim Biophys Acta* 1695 [1-3], Seite 225-33.
- [76] Qian, S. B.; Reits, E.; Neefjes, J.; Deslich, J. M.; Bennink, J. R. und Yewdell, J. W. (2006): Tight linkage between translation and MHC class I peptide ligand generation implies specialized antigen processing for defective ribosomal products, *J Immunol* 177 [1], Seite 227-33.
- [77] Amenta, J. S.; Sargus, M. J. und Baccino, F. M. (1978): Inhibition of basal protein degradation in rat embryo fibroblasts by cycloheximide: correlation with activities of lysosomal proteases, *J Cell Physiol* 97 [3 Pt 1], Seite 267-83.
- [78] Amshoff, C.; Jack, H. M. und Haas, I. G. (1999): Cycloheximide, a new tool to dissect specific steps in ER-associated degradation of different substrates, *Biol Chem* 380 [6], Seite 669-77.
- [79] Feldman, M. und Yagil, G. (1969): Does cycloheximide interfere with protein degradation?, *Biochem Biophys Res Commun* 37 [2], Seite 198-203.
- [80] Gerlinger, U. M.; Guckel, R.; Hoffmann, M.; Wolf, D. H. und Hilt, W. (1997): Yeast cycloheximide-resistant crl mutants are proteasome mutants defective in protein degradation, *Mol Biol Cell* 8 [12], Seite 2487-99.
- [81] Goldberg, A. L. und St John, A. C. (1976): Intracellular protein degradation in mammalian and bacterial cells: Part 2, *Annu Rev Biochem* 45, Seite 747-803.
- [82] Lee, H. K. und Marzella, L. (1994): Regulation of intracellular protein degradation with special reference to lysosomes: role in cell physiology and pathology, *Int Rev Exp Pathol* 35, Seite 39-147.
- [83] Wildenthal, K. und Griffin, E. E. (1976): Reduction by cycloheximide of lysosomal proteolytic enzyme activity and rate of protein degradation in organ-cultured hearts, *Biochim Biophys Acta* 444 [2], Seite 519-24.
- [84] Lyakhov, D. L.; He, B.; Zhang, X.; Studier, F. W.; Dunn, J. J. und McAllister, W. T. (1998): Pausing and termination by bacteriophage T7 RNA polymerase, *J Mol Biol* 280 [2], Seite 201-13.
- [85] Brandt, F.; Etchells, S. A.; Ortiz, J. O.; Elcock, A. H.; Hartl, F. U. und Baumeister, W. (2009): The native 3D organization of bacterial polysomes, *Cell* 136 [2], Seite 261-71.
- [86] Czaplinski, K.; Ruiz-Echevarria, M. J.; Gonzalez, C. I. und Peltz, S. W. (1999): Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover, *Bioessays* 21 [8], Seite 685-96.
- [87] Peters, B.; Bulik, S.; Tampe, R.; Van Endert, P. M. und Holzthutter, H. G. (2003): Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors, *J Immunol* 171 [4], Seite 1741-9.

- [88] Reits, E.; Griekspoor, A.; Neijssen, J.; Groothuis, T.; Jalink, K.; van Veelen, P.; Janssen, H.; Calafat, J.; Drijfhout, J. W. und Neefjes, J. (2003): Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I, *Immunity* 18 [1], Seite 97-108.
- [89] Esquivel, F.; Yewdell, J. und Bennink, J. (1992): RMA/S cells present endogenously synthesized cytosolic proteins to class I-restricted cytotoxic T lymphocytes, *J Exp Med* 175 [1], Seite 163-8.
- [90] Montoya, M. und Del Val, M. (1999): Intracellular rate-limiting steps in MHC class I antigen processing, *J Immunol* 163 [4], Seite 1914-22.
- [91] Villanueva, M. S.; Fischer, P.; Feen, K. und Pamer, E. G. (1994): Efficiency of MHC class I antigen processing: a quantitative analysis, *Immunity* 1 [6], Seite 479-89.
- [92] Vabulas, R. M. und Hartl, F. U. (2005): Protein synthesis upon acute nutrient restriction relies on proteasome function, *Science* 310 [5756], Seite 1960-3.
- [93] Drake, J. W. (1991): A constant rate of spontaneous mutation in DNA-based microbes, *Proc Natl Acad Sci U S A* 88 [16], Seite 7160-4.
- [94] Binder, R. J.; Blachere, N. E. und Srivastava, P. K. (2001): Heat shock protein-chaperoned peptides but not free peptides introduced into the cytosol are presented efficiently by major histocompatibility complex I molecules, *J Biol Chem* 276 [20], Seite 17163-71.
- [95] Esser, C.; Alberti, S. und Hohfeld, J. (2004): Cooperation of molecular chaperones with the ubiquitin/proteasome system, *Biochim Biophys Acta* 1695 [1-3], Seite 171-88.
- [96] Gullo, C. A. und Teoh, G. (2004): Heat shock proteins: to present or not, that is the question, *Immunol Lett* 94 [1-2], Seite 1-10.
- [97] Cascio, P.; Hilton, C.; Kisselev, A. F.; Rock, K. L. und Goldberg, A. L. (2001): 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide, *EMBO J* 20 [10], Seite 2357-66.
- [98] Bulik, S.; Peters, B.; Ebeling, C. und Holzhutter, H. (2004): Cytosolic processing of proteasomal cleavage products can enhance the presentation efficiency of MHC-1 epitopes, *Genome Inform* 15 [1], Seite 24-34.
- [99] Benaroudj, N.; Tarcsa, E.; Cascio, P. und Goldberg, A. L. (2001): The unfolding of substrates and ubiquitin-independent protein degradation by proteasomes, *Biochimie* 83 [3-4], Seite 311-8.
- [100] Yano, M.; Kanesaki, Y.; Koumoto, Y.; Inoue, M. und Kido, H. (2005): Chaperone activities of the 26S and 20S proteasome, *Curr Protein Pept Sci* 6 [2], Seite 197-203.
- [101] Poole, B. und Wibo, M. (1973): Protein degradation in cultured cells. The effect of fresh medium, fluoride, and iodoacetate on the digestion of cellular protein of rat fibroblasts, *J Biol Chem* 248 [17], Seite 6221-6.
- [102] May, R. M. (2004): Uses and abuses of mathematics in biology, *Science* 303 [5659], Seite 790-3.
- [103] Bulik, S.; Grimbs, S.; Huthmacher, C.; Selbig, J. und Holzhutter, H. G. (2009): Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws--a promising method for speeding up the kinetic modelling of complex metabolic networks, *FEBS J* 276 [2], Seite 410-24.
- [104] Grimbs, S.; Selbig, J.; Bulik, S.; Holzhutter, H. G. und Steuer, R. (2007): The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks, *Mol Syst Biol* 3, Seite 146.
- [105] Bulik, S.; Peters, B. und Holzhutter, H. G. (2005): Quantifying the contribution of defective ribosomal products to antigen production: a model-based computational analysis, *J Immunol* 175 [12], Seite 7957-64.

- [106] Rammensee, H.; Bachmann, J.; Emmerich, N. P.; Bachor, O. A. und Stevanovic, S. (1999): SYFPEITHI: database for MHC ligands and peptide motifs, *Immunogenetics* 50 [3-4], Seite 213-9.
- [107] Donnes, P. und Kohlbacher, O. (2005): Integrated modeling of the major events in the MHC class I antigen processing pathway, *Protein Sci* 14 [8], Seite 2132-40.
- [108] Nielsen, M.; Lundegaard, C.; Blicher, T.; Lamberth, K.; Harndahl, M.; Justesen, S.; Roder, G.; Peters, B.; Sette, A.; Lund, O. und Buus, S. (2007): NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence, *PLoS One* 2 [8], Seite e796.
- [109] Peters, B.; Bui, H. H.; Frankild, S.; Nielson, M.; Lundegaard, C.; Kostem, E.; Basch, D.; Lamberth, K.; Harndahl, M.; Fleri, W.; Wilson, S. S.; Sidney, J.; Lund, O.; Buus, S. und Sette, A. (2006): A community resource benchmarking predictions of peptide binding to MHC-I molecules, *PLoS Comput Biol* 2 [6], Seite e65.
- [110] Peters, B. und Sette, A. (2005): Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method, *BMC Bioinformatics* 6, Seite 132.
- [111] Peters, B.; Tong, W.; Sidney, J.; Sette, A. und Weng, Z. (2003): Examining the independent binding assumption for binding of peptide epitopes to MHC-I molecules, *Bioinformatics* 19 [14], Seite 1765-72.
- [112] Bradley, A. P. (1997): The use of the area under the roc curve in the evaluation of machine learning algorithms, *Pattern Recognition* 30 [7], Seite 1145-1159.
- [113] Udaka, K.; Wiesmuller, K. H.; Kienle, S.; Jung, G.; Tamamura, H.; Yamagishi, H.; Okumura, K.; Walden, P.; Suto, T. und Kawasaki, T. (2000): An automated prediction of MHC class I-binding peptides based on positional scanning with peptide libraries, *Immunogenetics* 51 [10], Seite 816-28.
- [114] IEDB Immune Epitope Database
- [115] Peters, B.; Sidney, J.; Bourne, P.; Bui, H. H.; Buus, S.; Doh, G.; Fleri, W.; Kronenberg, M.; Kubo, R.; Lund, O.; Nemazee, D.; Ponomarenko, J. V.; Sathiamurthy, M.; Schoenberger, S.; Stewart, S.; Surko, P.; Way, S.; Wilson, S. und Sette, A. (2005): The immune epitope database and analysis resource: from vision to blueprint, *PLoS Biol* 3 [3], Seite e91.
- [116] Gubler, B.; Daniel, S.; Armandola, E. A.; Hammer, J.; Caillat-Zucman, S. und van Endert, P. M. (1998): Substrate selection by transporters associated with antigen processing occurs during peptide binding to TAP, *Mol Immunol* 35 [8], Seite 427-33.
- [117] Uebel, S.; Kraas, W.; Kienle, S.; Wiesmuller, K. H.; Jung, G. und Tampe, R. (1997): Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries, *Proc Natl Acad Sci U S A* 94 [17], Seite 8976-81.
- [118] van Endert, P. M.; Riganelli, D.; Greco, G.; Fleischhauer, K.; Sidney, J.; Sette, A. und Bach, J. F. (1995): The peptide-binding motif for the human transporter associated with antigen processing, *J Exp Med* 182 [6], Seite 1883-95.
- [119] Van Endert, P. (2002): Affinitätsdaten für TAP von 9mer Peptiden, Bulik, S.
- [120] Van Endert, P. (2002): Affinitätsdaten von TAP zu längeren Peptiden, Bulik, S.
- [121] Tampe, R. (2002): Mix Matrix der TAP Affinität, Bulik, S.
- [122] Daniel, S.; Brusica, V.; Caillat-Zucman, S.; Petrovsky, N.; Harrison, L.; Riganelli, D.; Sinigaglia, F.; Gallazzi, F.; Hammer, J. und van Endert, P. M. (1998): Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules, *J Immunol* 161 [2], Seite 617-24.
- [123] Van Endert, P. (2002): Ala Matrix zur TAP Affinität von Peptiden, Bulik, S.
- [124] Uebel, S. und Tampe, R. (1999): Specificity of the proteasome and the TAP transporter, *Curr Opin Immunol* 11 [2], Seite 203-8.

- [125] van Endert, P. M.; Tampe, R.; Meyer, T. H.; Tisch, R.; Bach, J. F. und McDevitt, H. O. (1994): A sequential model for peptide binding and transport by the transporters associated with antigen processing, *Immunity* 1 [6], Seite 491-500.
- [126] Burgevin, A.; Saveanu, L.; Kim, Y.; Barilleau, E.; Kotturi, M.; Sette, A.; van Endert, P. und Peters, B. (2008): A detailed analysis of the murine TAP transporter substrate specificity, *PLoS One* 3 [6], Seite e2402.
- [127] Momburg, F.; Neefjes, J. J. und Hammerling, G. J. (1994): Peptide selection by MHC-encoded TAP transporters, *Curr Opin Immunol* 6 [1], Seite 32-7.
- [128] Brusic, V.; van Endert, P.; Zeleznikow, J.; Daniel, S.; Hammer, J. und Petrovsky, N. (1999): A neural network model approach to the study of human TAP transporter, *In Silico Biol* 1 [2], Seite 109-21.
- [129] Peters, B.; Janek, K.; Kuckelkorn, U. und Holzhutter, H. G. (2002): Assessment of proteasomal cleavage probabilities from kinetic analysis of time-dependent product formation, *J Mol Biol* 318 [3], Seite 847-62.
- [130] Toes, R. E.; Nussbaum, A. K.; Degermann, S.; Schirle, M.; Emmerich, N. P.; Kraft, M.; Laplace, C.; Zwinderman, A.; Dick, T. P.; Muller, J.; Schonfisch, B.; Schmid, C.; Fehling, H. J.; Stevanovic, S.; Rammensee, H. G. und Schild, H. (2001): Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products, *J Exp Med* 194 [1], Seite 1-12.
- [131] Emmerich, N. P.; Nussbaum, A. K.; Stevanovic, S.; Priemer, M.; Toes, R. E.; Rammensee, H. G. und Schild, H. (2000): The human 26 S and 20 S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate, *J Biol Chem* 275 [28], Seite 21140-8.
- [132] Tenzer, S.; Stoltze, L.; Schonfisch, B.; Dengjel, J.; Muller, M.; Stevanovic, S.; Rammensee, H. G. und Schild, H. (2004): Quantitative analysis of prion-protein degradation by constitutive and immuno-20S proteasomes indicates differences correlated with disease susceptibility, *J Immunol* 172 [2], Seite 1083-91.
- [133] Alvarez, I.; Sesma, L.; Marcilla, M.; Ramos, M.; Marti, M.; Camafeita, E. und de Castro, J. A. (2001): Identification of novel HLA-B27 ligands derived from polymorphic regions of its own or other class I molecules based on direct generation by 20 S proteasome, *J Biol Chem* 276 [35], Seite 32729-37.
- [134] Ayyoub, M.; Stevanovic, S.; Sahin, U.; Guillaume, P.; Servis, C.; Rimoldi, D.; Valmori, D.; Romero, P.; Cerottini, J. C.; Rammensee, H. G.; Pfreundschuh, M.; Speiser, D. und Levy, F. (2002): Proteasome-assisted identification of a SSX-2-derived epitope recognized by tumor-reactive CTL infiltrating metastatic melanoma, *J Immunol* 168 [4], Seite 1717-22.
- [135] Kessler, J. H.; Beekman, N. J.; Bres-Vloemans, S. A.; Verdijk, P.; van Veelen, P. A.; Kloosterman-Joosten, A. M.; Vissers, D. C.; ten Bosch, G. J.; Kester, M. G.; Sijts, A.; Wouter Drijfhout, J.; Ossendorp, F.; Offringa, R. und Melief, C. J. (2001): Efficient identification of novel HLA-A(*)0201-presented cytotoxic T lymphocyte epitopes in the widely expressed tumor antigen PRAME by proteasome-mediated digestion analysis, *J Exp Med* 193 [1], Seite 73-88.
- [136] Mommaas, B.; Kamp, J.; Drijfhout, J. W.; Beekman, N.; Ossendorp, F.; Van Veelen, P.; Den Haan, J.; Goulmy, E. und Mutis, T. (2002): Identification of a novel HLA-B60-restricted T cell epitope of the minor histocompatibility antigen HA-1 locus, *J Immunol* 169 [6], Seite 3131-6.
- [137] Morel, S.; Levy, F.; Burlet-Schiltz, O.; Brasseur, F.; Probst-Kepper, M.; Peitrequin, A. L.; Monsarrat, B.; Van Velthoven, R.; Cerottini, J. C.; Boon, T.; Gairin, J. E. und Van den Eynde, B. J. (2000): Processing of some antigens by the standard proteasome but not by the immunoproteasome results in poor presentation by dendritic cells, *Immunity* 12 [1], Seite 107-17.

- [138] Schultz, E. S.; Chapiro, J.; Lurquin, C.; Claverol, S.; Burlet-Schiltz, O.; Warnier, G.; Russo, V.; Morel, S.; Levy, F.; Boon, T.; Van den Eynde, B. J. und van der Bruggen, P. (2002): The production of a new MAGE-3 peptide presented to cytolytic T lymphocytes by HLA-B40 requires the immunoproteasome, *J Exp Med* 195 [4], Seite 391-9.
- [139] Sun, Y.; Sijts, A. J.; Song, M.; Janek, K.; Nussbaum, A. K.; Kral, S.; Schirle, M.; Stevanovic, S.; Paschen, A.; Schild, H.; Kloetzel, P. M. und Schadendorf, D. (2002): Expression of the proteasome activator PA28 rescues the presentation of a cytotoxic T lymphocyte epitope on melanoma cells, *Cancer Res* 62 [10], Seite 2875-82.
- [140] Tenzer, S.; Peters, B.; Bulik, S.; Schoor, O.; Lemmel, C.; Schatz, M. M.; Kloetzel, P. M.; Rammensee, H. G.; Schild, H. und Holzthutter, H. G. (2005): Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding, *Cell Mol Life Sci* 62 [9], Seite 1025-37.
- [141] Lucchiari-Hartz, M.; Lindo, V.; Hitziger, N.; Gaedicke, S.; Saveanu, L.; van Endert, P. M.; Greer, F.; Eichmann, K. und Niedermann, G. (2003): Differential proteasomal processing of hydrophobic and hydrophilic protein regions: contribution to cytotoxic T lymphocyte epitope clustering in HIV-1-Nef, *Proc Natl Acad Sci U S A* 100 [13], Seite 7755-60.
- [142] Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J. und Wheeler, D. L. (2008): GenBank, *Nucleic Acids Res* 36 [Database issue], Seite D25-30.
- [143] Holzthutter, H. G.; Frommel, C. und Kloetzel, P. M. (1999): A theoretical approach towards the identification of cleavage-determining amino acid motifs of the 20 S proteasome, *J Mol Biol* 286 [4], Seite 1251-65.
- [144] Holzthutter, H. G. und Kloetzel, P. M. (2000): A kinetic model of vertebrate 20S proteasome accounting for the generation of major proteolytic fragments from oligomeric peptide substrates, *Biophys J* 79 [3], Seite 1196-205.
- [145] Kesmir, C.; Nussbaum, A. K.; Schild, H.; Detours, V. und Brunak, S. (2002): Prediction of proteasome cleavage motifs by neural networks, *Protein Eng* 15 [4], Seite 287-96.
- [146] Kuttler, C.; Nussbaum, A. K.; Dick, T. P.; Rammensee, H. G.; Schild, H. und Hader, K. P. (2000): An algorithm for the prediction of proteasomal cleavages, *J Mol Biol* 298 [3], Seite 417-29.
- [147] Nussbaum, A. K.; Kuttler, C.; Hader, K. P.; Rammensee, H. G. und Schild, H. (2001): PAPROC: a prediction algorithm for proteasomal cleavages available on the WWW, *Immunogenetics* 53 [2], Seite 87-94.
- [148] Hader, K. P.; Kuttler, C. und Nussbaum, A. K. (2004): Cleaving proteins for the immune system, *Math Biosci* 188, Seite 63-79.
- [149] Kloetzel, P. M. (2001): Antigen processing by the proteasome, *Nat Rev Mol Cell Biol* 2 [3], Seite 179-87.
- [150] Kawahara, M.; York, I. A.; Hearn, A.; Farfan, D. und Rock, K. L. (2009): Analysis of the role of tripeptidyl peptidase II in MHC class I antigen presentation in vivo, *J Immunol* 183 [10], Seite 6069-77.
- [151] Reits, E.; Neijssen, J.; Herberts, C.; Benckhuijsen, W.; Janssen, L.; Drijfhout, J. W. und Neefjes, J. (2004): A major role for TPPII in trimming proteasomal degradation products for MHC class I antigen presentation, *Immunity* 20 [4], Seite 495-506.
- [152] Seifert, U.; Maranon, C.; Shmueli, A.; Desoutter, J. F.; Wesoloski, L.; Janek, K.; Henklein, P.; Diescher, S.; Andrieu, M.; de la Salle, H.; Weinschenk, T.; Schild, H.; Laderach, D.; Galy, A.; Haas, G.; Kloetzel, P. M.; Reiss, Y. und Hosmalin, A. (2003): An essential role for tripeptidyl peptidase in the generation of an MHC class I epitope, *Nat Immunol* 4 [4], Seite 375-9.

- [153] Altuvia, Y. und Margalit, H. (2000): Sequence signals for generation of antigenic peptides by the proteasome: implications for proteasomal cleavage mechanism, *J Mol Biol* 295 [4], Seite 879-90.
- [154] Schatz, M. M.; Peters, B.; Akkad, N.; Ullrich, N.; Martinez, A. N.; Carroll, O.; Bulik, S.; Rammensee, H. G.; van Endert, P.; Holzhutter, H. G.; Tenzer, S. und Schild, H. (2008): Characterizing the N-terminal processing motif of MHC class I ligands, *J Immunol* 180 [5], Seite 3210-7.
- [155] Kondo, A.; Sidney, J.; Southwood, S.; del Guercio, M. F.; Appella, E.; Sakamoto, H.; Celis, E.; Grey, H. M.; Chesnut, R. W.; Kubo, R. T. und Sette, A. (1995): Prominent roles of secondary anchor residues in peptide binding to HLA-A24 human class I molecules, *J Immunol* 155 [9], Seite 4307-12.
- [156] Kondo, A.; Sidney, J.; Southwood, S.; del Guercio, M. F.; Appella, E.; Sakamoto, H.; Grey, H. M.; Celis, E.; Chesnut, R. W.; Kubo, R. T. und Sette, A. (1997): Two distinct HLA-A*0101-specific submotifs illustrate alternative peptide binding modes, *Immunogenetics* 45 [4], Seite 249-58.
- [157] Sidney, J.; Grey, H. M.; Southwood, S.; Celis, E.; Wentworth, P. A.; del Guercio, M. F.; Kubo, R. T.; Chesnut, R. W. und Sette, A. (1996): Definition of an HLA-A3-like supermotif demonstrates the overlapping peptide-binding repertoires of common HLA molecules, *Hum Immunol* 45 [2], Seite 79-93.
- [158] Sidney, J.; Southwood, S.; del Guercio, M. F.; Grey, H. M.; Chesnut, R. W.; Kubo, R. T. und Sette, A. (1996): Specificity and degeneracy in peptide binding to HLA-B7-like class I molecules, *J Immunol* 157 [8], Seite 3480-90.
- [159] Sidney, J.; Southwood, S.; Pasquetto, V. und Sette, A. (2003): Simultaneous prediction of binding capacity for multiple molecules of the HLA B44 supertype, *J Immunol* 171 [11], Seite 5964-74.
- [160] Parker, K. C.; Bednarek, M. A. und Coligan, J. E. (1994): Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains, *J Immunol* 152 [1], Seite 163-75.
- [161] Python Offizielle Internetseite der Python Community
- [162] Vita, R.; Zarebski, L.; Greenbaum, J. A.; Emami, H.; Hoof, I.; Salimi, N.; Damle, R.; Sette, A. und Peters, B. The immune epitope database 2.0, *Nucleic Acids Res* 38 [Database issue], Seite D854-62.
- [163] Hearn, A.; York, I. A. und Rock, K. L. (2009): The specificity of trimming of MHC class I-presented peptides in the endoplasmic reticulum, *J Immunol* 183 [9], Seite 5526-36.
- [164] Kisselev, A. F.; Akopian, T. N.; Woo, K. M. und Goldberg, A. L. (1999): The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation, *J Biol Chem* 274 [6], Seite 3363-71.
- [165] Mishto, M.; Luciani, F.; Holzhutter, H. G.; Bellavista, E.; Santoro, A.; Textoris-Taube, K.; Franceschi, C.; Kloetzel, P. M. und Zaikin, A. (2008): Modeling the in vitro 20S proteasome activity: the effect of PA28- α and of the sequence and length of polypeptides on the degradation kinetics, *J Mol Biol* 377 [5], Seite 1607-17.
- [166] Gillespie, Daniel T. (1976): A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of Computational Physics* 22 [4], Seite 403-434.
- [167] ExPASy ExPASy Proteomics Server
- [168] Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D. und Bairoch, A. (2003): ExPASy: The proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res* 31 [13], Seite 3784-8.

- [169] Goldobin, D. S. und Zaikin, A. (2009): Towards quantitative prediction of proteasomal digestion patterns of proteins, *Journal of Statistical Mechanics-Theory and Experiment*, Seite -.
- [170] Kisselev, A. F.; Kaganovich, D. und Goldberg, A. L. (2002): Binding of hydrophobic peptides to several non-catalytic sites promotes peptide hydrolysis by all active sites of 20 S proteasomes. Evidence for peptide-induced channel opening in the alpha-rings, *J Biol Chem* 277 [25], Seite 22260-70.
- [171] Kloss, A.; Henklein, P.; Siele, D.; Schmolke, M.; Apcher, S.; Kuehn, L.; Sheppard, P. W. und Dahlmann, B. (2009): The cell-penetrating peptide octa-arginine is a potent inhibitor of proteasome activities, *Eur J Pharm Biopharm* 72 [1], Seite 219-25.
- [172] Schmidtke, G.; Emch, S.; Groettrup, M. und Holzhutter, H. G. (2000): Evidence for the existence of a non-catalytic modifier site of peptide hydrolysis by the 20 S proteasome, *J Biol Chem* 275 [29], Seite 22056-63.
- [173] Bech-Otschir, D.; Helfrich, A.; Enenkel, C.; Consiglieri, G.; Seeger, M.; Holzhutter, H. G.; Dahlmann, B. und Kloetzel, P. M. (2009): Polyubiquitin substrates allosterically activate their own degradation by the 26S proteasome, *Nat Struct Mol Biol* 16 [2], Seite 219-25.
- [174] Knuehl, C.; Spee, P.; Ruppert, T.; Kuckelkorn, U.; Henklein, P.; Neefjes, J. und Kloetzel, P. M. (2001): The murine cytomegalovirus pp89 immunodominant H-2Ld epitope is generated and translocated into the endoplasmic reticulum as an 11-mer precursor peptide, *J Immunol* 167 [3], Seite 1515-21.
- [175] Yang, O. O.; Kalams, S. A.; Rosenzweig, M.; Trocha, A.; Jones, N.; Koziel, M.; Walker, B. D. und Johnson, R. P. (1996): Efficient lysis of human immunodeficiency virus type 1-infected cells by cytotoxic T lymphocytes, *J Virol* 70 [9], Seite 5799-806.
- [176] Scarlata, S. und Carter, C. (2003): Role of HIV-1 Gag domains in viral assembly, *Biochim Biophys Acta* 1614 [1], Seite 62-72.
- [177] Lauemoller, S. L.; Holm, A.; Hilden, J.; Brunak, S.; Holst Nissen, M.; Stryhn, A.; Ostergaard Pedersen, L. und Buus, S. (2001): Quantitative predictions of peptide binding to MHC class I molecules using specificity matrices and anchor-stratified calibrations, *Tissue Antigens* 57 [5], Seite 405-14.
- [178] Uebel, S.; Wiesmuller, K. H.; Jung, G. und Tampe, R. (1999): Peptide libraries in cellular immune recognition, *Curr Top Microbiol Immunol* 243, Seite 1-21.
- [179] Mitchell, Tom M. (1997): *Machine learning*, McGraw-Hill, New York ISBN: 0-07-115467-10-07-042807-7978-0-07-115467-3.
- [180] Bhasin, M. und Raghava, G. P. (2004): Prediction of CTL epitopes using QM, SVM and ANN techniques, *Vaccine* 22 [23-24], Seite 3195-204.
- [181] Eddy, S. R. (2004): What is a hidden Markov model?, *Nat Biotechnol* 22 [10], Seite 1315-6.
- [182] Mukherjee, S. und Mitra, S. (2005): Hidden Markov Models, grammars, and biology: a tutorial, *J Bioinform Comput Biol* 3 [2], Seite 491-526.
- [183] Schuster-Bockler, B. und Bateman, A. (2007): An introduction to hidden Markov models, *Curr Protoc Bioinformatics Appendix* 3, Seite Appendix 3A.
- [184] Thorvaldsen, S. (2005): A tutorial on Markov models based on Mendel's classical experiments, *J Bioinform Comput Biol* 3 [6], Seite 1441-60.
- [185] Udaka, K.; Mamitsuka, H.; Nakaseko, Y. und Abe, N. (2002): Empirical evaluation of a dynamic experiment design method for prediction of MHC class I-binding peptides, *J Immunol* 169 [10], Seite 5744-53.
- [186] Zhang, C.; Bickis, M. G.; Wu, F. X. und Kusalik, A. J. (2006): Optimally-connected hidden markov models for predicting MHC-binding peptides, *J Bioinform Comput Biol* 4 [5], Seite 959-80.

- [187] Noble, W. S. (2006): What is a support vector machine?, Nat Biotechnol 24 [12], Seite 1565-7.
- [188] Bhasin, M.; Lata, S. und Raghava, G. P. (2007): TAPPred prediction of TAP-binding peptides in antigens, Methods Mol Biol 409, Seite 381-6.
- [189] Riedesel, H.; Kolbeck, B.; Schmetzer, O. und Knapp, E. W. (2004): Peptide binding at class I major histocompatibility complex scored with linear functions and support vector machines, Genome Inform 15 [1], Seite 198-212.
- [190] Breiman, Leo (1998): Classification and regression trees, 1. repr. Auflage, Chapman & Hall, Boca Raton, ISBN: 0-412-04841-8.
- [191] Breiman, L. (2001): Random forests, Machine Learning 45 [1], Seite 5-32.
- [192] Hansen, L.; Lee, E. A.; Hestir, K.; Williams, L. T. und Farrelly, D. (2009): Controlling feature selection in random forests of decision trees using a genetic algorithm: classification of class I MHC peptides, Comb Chem High Throughput Screen 12 [5], Seite 514-9.

Publikationen

Bulik, S., S. Grimbs, et al. (2009). "Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws--a promising method for speeding up the kinetic modelling of complex metabolic networks." *FEBS J* 276(2): 410-24.

Bulik, S., B. Peters, et al. (2004). "Cytosolic processing of proteasomal cleavage products can enhance the presentation efficiency of MHC-1 epitopes." *Genome Inform* 15(1): 24-34.

Bulik, S., B. Peters, et al. (2005). "Quantifying the contribution of defective ribosomal products to antigen production: a model-based computational analysis." *J Immunol* 175(12): 7957-64.

Dandekar, T., F. Moldenhauer, et al. (2003). "A method for classifying metabolites in topological pathway analyses based on minimization of pathway number." *Biosystems* 70(3): 255-70.

Gille, C., C. Bolling, et al. (2010). "HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology." *Mol Syst Biol* 6: 411.

Grimbs, S., J. Selbig, et al. (2007). "The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks." *Mol Syst Biol* 3: 146.

Huthmacher, C., A. Hoppe, et al. (2010). "Antimalarial drug targets in *Plasmodium falciparum* predicted by stage-specific metabolic network analysis." *BMC Syst Biol* 4: 120.

Peters, B., S. Bulik, et al. (2003). "Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors." *J Immunol* 171(4): 1741-9.

Rother, K., S. Hoffmann, et al. (2010). "IGERS: inferring Gibbs energy changes of biochemical reactions from reaction similarities." *Biophys J* 98(11): 2478-86.

Schatz, M. M., B. Peters, et al. (2008). "Characterizing the N-terminal processing motif of MHC class I ligands." *J Immunol* 180(5): 3210-7.

Tenzer, S., B. Peters, et al. (2005). "Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding." *Cell Mol Life Sci* 62(9): 1025-37.

8 Anhang

8.1 Die stabilisierte Matrix Methode (SMM)[110]

Die Repräsentation der Affinität aller möglichen Peptide¹¹ der Länge L mit einer $L \times 20$ Matrix ist möglich, wenn man annimmt, dass jede Aminosäure in einem Peptid unabhängig von den anderen zur Bindung an das MHC Molekül beiträgt.

Aus einer Matrix ergibt sich die Affinität A eines Peptides k durch:

(25)

$$A_k = P_0 + \sum_{i=1}^L P_{i,a_k(i)}$$

$a_k(i)$ ist die Aminosäure des Peptides k , die sich an der Position i befindet. Man erhält die Einträge der Matrix $P_{i,a}$ und den konstanten Summanden P_0 indem man den Abstand

(26)

$$\Phi(\{P_{i,a}\}, P_0) = \sum_k \|Y_k - A_k\|$$

zwischen den von der Matrix vorhergesagten und den gemessenen Werten (Y_k) eines Trainingsdatensatzes minimiert. Hat man dieses durchgeführt, besitzt man eine Matrix, die den Trainingsdatensatz optimal beschreibt, aber möglicherweise unabhängige Datensätze schlechter wiedergibt, da in Experimenten die Ergebnisse immer verrauscht sind und man die Parameter an das Rauschen im Trainingsdatensatz mit anpasst (overfitting).

Zur Reduzierung des Overfittings wird in die Abstandsfunktion (Gleichung 26) ein weiterer Term eingeführt. Dieser Term bewirkt, dass die Beträge der einzelnen Matrixeinträge $P_{i,a}$ so klein wie möglich werden, wobei ein Kompromiss zwischen guter Reproduktion des Trainingsdatensatzes und kleinen Matrixeinträgen gefunden werden muss. Wenn also für einzelne Parameter keine oder nur wenig experimentelle Information zur Verfügung steht, wird der Wert klein, sind viele Informationen vorhanden, gibt es einen dementsprechenden Wert in der Matrix.

¹¹ Die Anzahl aller möglichen Peptide der Länge L ergibt sich nach 20^L , wobei 20 die Anzahl der proteinogenen Aminosäuren ist. Für 9mer Peptide ist $20^9 = 5.12 \cdot 10^{11}$.

(27)

$$\Psi(\{P_{i,a}\}, P_0, \lambda) = \Phi(\{P_{i,a}\}, P_0) + \lambda \sum_{i,a} (P_{i,a})^2$$

Für das Training der Matrixeinträge muss der neue Abstand Ψ minimiert werden. Das optimale λ wird bestimmt, indem man λ systematisch variiert und den Abstand Ψ mittels 10-facher Kreuzvalidierung¹² berechnet. Der Abstand ist minimal für das optimale λ .

Man erhält so eine Matrix, die weniger stark „verrauscht“ ist und dafür den Trainingsdatensatz nicht optimal beschreibt. Diese stabilisierte Matrix Methode ist für die Affinitätsvorhersage von Peptiden entwickelt worden, lässt sich aber auch auf andere Problemstellungen anwenden.

Voraussetzung für den Erfolg aller Matrixmethoden ist jedoch, dass die Annahme der unabhängigen Bindung in guter Näherung erfüllt ist.

8.2 Weitere Methoden zur Vorhersage von Peptideigenschaften (Übersicht)

Hier soll eine kurze Übersicht über eine Reihe von weiteren Methoden gegeben werden. Diese Zusammenstellung erhebt keinen Anspruch auf Vollständigkeit und soll nur ein paar Alternativen zu den hier verwendeten Methoden für die Erstellung von statistischen Modellen vorstellen.

8.2.1 Direktes Messen der Affinität mittels Peptidbibliotheken

Die einzelnen Beiträge der Aminosäuren in der Affinitätsmatrix lassen sich mittels Peptidbibliotheken elegant messen. Eine Peptidbibliothek besteht aus 9mer Peptiden, die alle an einer Stelle eine feste Aminosäure haben (zum Beispiel an der vierten Position ein Leucin) und an den anderen Positionen eine Mischung aller Aminosäuren. Die Bibliothek wird dann als XXXLXXXXX angegeben, wobei X für eine Mischung aus Aminosäuren steht. So haben 5 Prozent der Peptide in einer Bibliothek an einer mit X bezeichneten Stelle ein Alanin, 5 Prozent ein Cystein und so weiter. Aus technischen Gründen wird in der Praxis kein Cystein eingesetzt, da bei den benötigten Konzentrationen Schwefelbrücken gebildet werden, die zu verfälschten Daten führen können. Auch eine Reduktion der Anzahl der

¹² Für die n-fache Kreuzvalidierung wird der Trainingsdatensatz in n gleichgroße Teilmengen zerlegt. Dann wird jede Teilmenge (=Blindset) einmal aus dem Gesamttrainingsdatensatz entfernt und das Modell trainiert. Anschließend wird das Modell auf das Blindset angewendet, um die Performance des Modells zu bewerten.

verbleibenden 19 Aminosäuren auf einen repräsentativen Satz von 6 Aminosäuren ist aus technischen Gründen mitunter sinnvoll. Durch diese technischen Modifikationen sind nicht mehr 5 Prozent jeder Aminosäure an einer mit X bezeichneten Position vertreten, sondern der Anteil der bei der Synthese des Peptids eingesetzten Menge an dieser Aminosäure.

Man erhält so insgesamt 171 Peptidbibliotheken, wenn man an jeder Position jede der 20 proteinogenen Aminosäuren außer Cystein einsetzt. Die Affinität ist gegen eine komplett randomisierte, das heißt an jeder Stelle aus einer Mischung von Aminosäure bestehende, Peptidsequenz angegeben. Auf diese Weise erhält man den Einfluss der Aminosäure an der jeweiligen Position auf die Affinität des Gesamtpeptides und somit direkt den entsprechenden Eintrag in der Affinitätsmatrix.

Diese Methode zur Affinitätsbestimmung ist schon erfolgreich auf verschiedenen Gebieten eingesetzt worden[177, 178].

8.2.2 Das Einbeziehen von Paarkoeffizienten

Die Annahme unabhängiger Beiträge zur Affinität der Peptide kann beispielsweise durch Paarkoeffizienten erweitert werden. Paarkoeffizienten beschreiben die Änderung der Affinität beim gleichzeitigen Auftreten zweier Aminosäuren an bestimmten Positionen des Peptids. Man erhält für 9mer Peptide $9 \cdot 8/2 \cdot 20^2 = 14400$ Paarkoeffizienten und es ist nicht möglich, Aussagen über alle Paarkoeffizienten aus einer geringen Anzahl an Beobachtungen zu gewinnen.

Zum Bestimmen der Paarkoeffizienten wird folgendermaßen vorgegangen: Der Startpunkt ist eine Bewertungsmatrix. Diese wird beim Training der Paarkoeffizienten nicht verändert. Dann werden Kandidaten für Paarkoeffizienten gesucht, das heißt es werden nur Koeffizienten für die Paare bestimmt, für die im Datensatz mindestens 10 Beobachtungen vorhanden sind. Das Training der Paarkoeffizienten erfolgt vollkommen analog zum Training der SMM Matrix.

(28)

$$A'_k = A_k + \sum_{i=1}^L \sum_{i'=1; i' \neq i}^L P'_{i, a_k(i), i', a_k(i')}$$

$$\Phi'(\{P_{i,a}\}, \{P'_{i, a_k(i), i', a_k(i')}\}, P_0) = \sum_k \|Y_k - A'_k\|$$

$$\Psi'(\{P_{i,a}\}, P_0, \lambda) = \Phi'(\{P_{i,a}\}, \{P'_{i, a_k(i), i', a_k(i')}\}, P_0) + \lambda \sum_{i, i'}^L (P'_{i, a_k(i), i', a_k(i')})^2$$

Die P' sind die Paarkoeffizienten zwischen den Aminosäuren a und a' des Peptides k an den Positionen i und i' . Der regulierende Parameter λ' erfüllt die gleiche Funktion beim Training der Paarkoeffizienten wie der Parameter λ beim Training der Matrixeinträge der SMM Matrix.

Es werden die Paarkoeffizienten vernachlässigt, die auf einem Teil der 10 Trainingsdatensätze ein positives und auf einem anderen Teil ein negatives Vorzeichen haben, da über diese Paarkoeffizienten entweder nicht genügend Informationen vorhanden sind, um sie exakt zu bestimmen, oder diese Paarkoeffizienten so nahe an Null liegen, dass sie auf unterschiedlichen Teilmengen des Datensatzes unterschiedliche Vorzeichen annehmen. Die verbleibenden Paarkoeffizienten werden dann analog zur stabilisierten Matrix Methode mit systematisch variiertem λ' trainiert[111].

Die für die Paarkoeffizienten als Grundlage verwendete Matrix muss nicht unbedingt mit der SMM Methode erstellt worden sein.

Wenn durch Paarkoeffizienten eine deutliche Verbesserung der Vorhersagequalität erreicht werden kann, ist das ein Hinweis darauf, dass die Annahme unabhängiger Beiträge zur Affinität nicht gerechtfertigt war beziehungsweise mit dem resultierenden Modell nicht die wesentlichen Eigenschaften des Prozesses erfasst werden. Durch das Einführen vieler neuer Parameter ist jedoch eine gewisse Verbesserung zu erwarten und rigoroses verringern von Overfitting notwendig.

8.2.3 Künstliche neuronale Netze, Hidden Markov Modells, Regression Trees und Random Forests

Viele Methoden des maschinellen Lernens wurden und werden auf dem Gebiet der Antigenvorhersage eingesetzt. Maßgeblich für die Eignung dieser Methoden ist zum einen der Erfolg, den sie bei der Vorhersage unabhängiger Daten haben. Andererseits ist es in dieser Arbeit nicht das Primärziel eine möglichst hohe Qualität der Vorhersage zu erreichen, sondern vor allem die zugrunde liegenden Eigenschaften des Systems zu beschreiben und zu verstehen. Für die meisten Methoden des maschinellen Lernens gibt es komfortable Implementationen der benötigten Algorithmen, jedoch ist das Erkennen von Systemeigenschaften oft eingeschränkt.

Künstliche neuronale Netzwerke sind Verschaltungen von künstlichen Neuronen, die Eingangssignale wichten, umwandeln, weiterleiten und zu einer Ausgabe zusammenfassen. Es gibt unterschiedliche Arten von Neuronen (Perzeptron, lineare oder sigmoidale Einheit) und unterschiedliche Verschaltungen (vorwärts gerichtet oder zyklisch) zwischen ihnen.

Die Komplexität, also die Anzahl der Neuronen und Schichten des Netzwerks, muss dem Problem sowie der Größe des Trainingsdatensatzes angemessen sein[179]. Die internen Gewichte werden durch „Backpropagation“¹³ trainiert und sind in der Regel nicht mehr mit mechanistischen Eigenschaften des betrachteten Systems zu identifizieren. Um Overfitting zu vermeiden oder zumindest zu reduzieren, wird jedes Gewicht bei jeder Iteration des Trainings leicht reduziert (weight decay). Dieses ist Äquivalent zum Einfügen eines Strafterms zur Gewichtshöhe in die Zielfunktion der Optimierung. Neuronale Netze wurden auf eine Vielzahl von biologischen Systemen angewendet. Für die Antigenpräsentation wurden beispielsweise CTL Epitope vorhergesagt oder auch die Bindung von Peptiden an TAP[128, 180]. Vom Prinzip her entspricht die zuvor beschriebene Matrixmethode einem einschichtigen neuronalen Netz, bei dem die entsprechenden Gewichte trainiert werden. Die Reduktion des Overfittings durch λ (Gleichung 27) entspricht dem „weight decay“ beim Training des neuronalen Netzwerks. Paarkoeffizienten würde man mit einem zweischichtigen Neuronalen Netz erfassen können, wobei hier das Netz schon komplexer ist, als nur die 14400 + 180 Parameter der kombinierten Matrix und Paarkoeffizienten Methode.

Hidden Markov Modelle sind eine Menge beziehungsweise Kette ‚versteckter‘ interner Zustände, die durch Übergangswahrscheinlichkeiten verknüpft sind. Beim durchlaufen der Markov-Kette von einem Start zu einem Endzustand werden verschiedene innere Zustände durchlaufen, die eine beobachtbare Sequenz erzeugen[181, 182, 183, 184]. Hidden Markov Modelle werden prinzipiell für verschiedene Fragestellungen verwendet: i) Aus der beobachtbaren Sequenz soll die wahrscheinliche Reihenfolge der internen Zustände geschlussfolgert werden. ii) Es soll diskriminiert werden, welche Realisierung eines Hidden Markov Modells am wahrscheinlichsten die beobachteten Sequenzen generieren würde. iii) Welcher Endzustand ist am wahrscheinlichsten, wenn man die beobachtete Sequenz zugrunde legt. Die dritte Fragestellung ist die für die Antigenvorhersage verwendete. Hidden Markov Modelle wurden bereits für die Bindung von Peptiden an MHC Moleküle verwendet[185, 186].

Support Vector Machine (SVM) bezeichnet einen Klassifizierungsalgorithmus, der mittels einer Hyperebene die Trainingsdaten in zwei Gruppen unterteilt. Eine feinere Unterteilung in mehr Klassen ist durch die Einführung weiterer Hyperebenen möglich[187]. Die Methode wurde für die Bindung von Peptiden an TAP und MHC angewendet[188, 189].

¹³ Backpropagation bezeichnet ein Gradientenverfahren zur Optimierung der Gewichte des neuronalen Netzwerks. Hierbei wird der Fehler (= Differenz zwischen Netzwerkausgabe und gewünschtem Ergebnis) beim nächsten Iterationsschritt zur Eingabe hinzugefügt.

Regressionsbäume (Regression Trees) stellen ebenfalls eine Klassifizierungsmethode dar. Ein Regressionsbaum besteht aus einer Wurzel, einer beliebig großen Anzahl von inneren Knoten beziehungsweise Verzweigungen und den als Blätter bezeichneten Klassifizierungen. Jeder Knoten inklusive der Wurzel repräsentiert eine logische Regel und führt zu einer binären Entscheidung, die entweder zu einem weiteren Knoten oder zu einem Blatt führt. Beim Aufstellen des Modells müssen die logischen Regeln gefunden werden, die am geeignetsten sind, den Trainingsdatensatz zu klassifizieren[190]. Ein Vorteil der Methode ist, dass man an den erstellten Regeln lernen kann, nach welchen Gesichtspunkten ein bestimmtes Merkmal auftritt. Auch diese Methode wurde verwendet um MHC Molekül Binder vorherzusagen[111].

Random Forests stellen eine Erweiterung der Regressionsbäume dar. Es werden viele zufällig unterschiedene Regressionsbäume gebildet, mit denen dann jeweils die Klassifizierung durchgeführt wird. Die endgültige Klassifizierung ist dann entweder die mehrheitliche Klassifizierung der einzelnen Regressionsbäume oder eine anteilige Klassifizierung, dass heißt eine Wahrscheinlichkeit zu einer Gruppe zu gehören[191]. Vorteil dieser Methode ist die im Algorithmus implementierte Reduktion von Overfitting. Hansen et al. haben mit dieser Methode MHC Molekül Bindung vorhergesagt[192].

Klassifizierungsmethoden beinhalten den Nachteil, dass nur begrenzt quantitative Aussagen möglich sind. Für die hier betrachteten Fragen ist es nicht nur wichtig zu wissen, ob ein Molekül bindet, sondern auch wie stark, wozu man annähernd kontinuierliche Ergebnisse braucht.

8.2.4 Supervised learning

Die im Trainingsdatensatz enthaltene Information für die Modellparameter ist normalerweise nicht gleichmäßig verteilt. Daher sind die Werte einzelner Parameter sicherer beziehungsweise weniger stark mit einem Fehler belastet als andere. Supervised learning (überwachtes Lernen) ist eine Methode die daran anknüpft und die größten Fehler beziehungsweise Unsicherheiten in den Parametern mittels gezielter Erweiterung des Trainingsdatensatzes beseitigt.

Aus dem ursprünglichen Trainingsdatensatz werden viele neue Datensätze durch bootstrapping with replacement¹⁴ erstellt. Daraufhin wird das Modell jeweils mit einem neuen Datensatz trainiert und dazu verwendet für sehr viele verschiedene zufällig generierte Sequenzen Vorhersagen zu machen. Die Sequenzen, für die die Modelle die unterschiedlichsten Werte vorhersagen, sind offensichtlich Sequenzen, über die man noch nicht genug Informationen hat. Diese werden synthetisiert und gemessen, und man erhält eine erweiterte Datenbasis. Diese Schritte können mehrmals wiederholt werden und man erhält ein Modell, das besser sein sollte, als ein Modell, das auf einer zufällig ausgewählten oder einseitigen Datenbasis beruht. Udaka et al. Verwendeten diese Methode, um ein Hidden Markov Modell der Affinität von MHC I Molekülen zu verbessern[185].

¹⁴ Bootstrapping with replacement: Diese Methode eignet sich dafür, zu prüfen, wie stark der Einfluss einzelner Datenpunkte auf den gesamten Datensatz ist bzw. wie stark die einzelnen Datenpunkte Parameterwerte bestimmen, die auf dem Datensatz trainiert werden. Als Ausgangspunkt hat man einen Datensatz mit n Werten. Aus diesem Datensatz zieht man n mal zufällig einen Wert, ohne dass der Ausgangsdatsatz verändert wird. Die n gezogenen Werte bilden einen neuen Datensatz, der durch „bootstrapping with replacement“ erstellt worden ist.

Im Allgemeinen wird der neue Datensatz einige Datenpunkte des ursprünglichen Datensatzes mehrfach enthalten und andere dafür nicht.

Erklärung zur Selbstständigkeit

Ich erkläre, dass ich die hier dargestellten Ergebnisse Selbstständig und ohne weitere als die aufgeführten Hilfsmittel erbracht habe.

Die Arbeiten zu TAP, proteasomaler Schnittvorhersage und N-terminalem Trimmen wurden in Zusammenarbeit mit Björn Peters durchgeführt.

Die experimentellen Ergebnisse wurden von entsprechenden Partnern erzeugt. An den entsprechenden Stellen dieser Arbeit wird darauf hingewiesen.